

FT Protein 400.575  
 /note= "Zeta membrane spanning and intracellular domain"  
 XX  
 XX MO9215322-A1.  
 XX  
 XX 17-SEP-1992.  
 XX  
 XX 06-MAR-1992; 92WO-US001785.  
 XX  
 XX 07-MAR-1991; 91US-00665961.  
 XX  
 XX (GEHO ) GEN HOSPITAL CORP.  
 XX  
 XX Seed B, Romeo C, Kolanus W;  
 XX  
 XX WPI: 1992-331474/40.  
 XX N-PSDB; AAQ28704.  
 XX  
 XX Therapeutic cells expressing chimeric receptors - directing cellular  
 PT response to an infective agent, useful in treating HIV-1, AIDS  
 PT Pneumocystis carinii infections etc.  
 XX  
 XX Example 2; Page 72-73; 114pp; English.  
 XX  
 XX This sequence represents a fusion protein between the CD4 extracellular  
 CC domain and the zeta protein membrane spanning domain and intracellular  
 CC domain. Zeta is a 32 kD type I integral membrane homodimer which has a 9  
 CC residue extracellular domain and a 112/113 residue intracellular domain  
 CC for mouse and human protein respectively. In the production of the CD4  
 CC receptor chimera, the zeta cDNA was isolated from the HBB-A17 tumour cell  
 CC line and from human natural killer cells. The zeta cDNA was joined to the  
 CC extracellular domain of an engineered form of CD4 possessing a BamHI site  
 CC just upstream of the membrane spanning domain, by a BamI site naturally  
 CC present a few residues upstream of the membrane spanning domain. (Updated  
 CC on 25-MAR-2003 to correct FN field.)  
 XX  
 XX Sequence 575 AA;  
 SQ

Query Match 59.7%; Score 2036.5; DB 2; Length 575;  
 Best Local Similarity 98.3%; Pred. No. 66-102;  
 Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MNRGVPFRHLLVQLALPAATQGNKVLGKGGDTVELCTASQKSIQPHMKNNOIK 60  
 DB 1 MNRGVPFRHLLVQLALPAATQGNKVLGKGGDTVELCTASQKSIQPHMKNNOIK 60  
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120  
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120  
 QY 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
 DB 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
 QY 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
 DB 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
 QY 181 TWCTCTVQONQKVEFKIDIVLAFOKASSIYKKEGEOVESFPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTCTVQONQKVEFKIDIVLAFOKASSIYKKEGEOVESFPLAFTVEKLTGSGELMW 240  
 QY 241 QABRASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKLPJHLTLPOALPOYAGSGNITLA 300  
 DB 241 QABRASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKLPJHLTLPOALPOYAGSGNITLA 300  
 QY 301 LEATGTGLHGEVNLVVMRATQLOKNTCEVWGPTSPKLMSTLKENKEAKVSRKEKRVWV 360  
 DB 301 LEATGTGLHGEVNLVVMRATQLOKNTCEVWGPTSPKLMSTLKENKEAKVSRKEKRVWV 360  
 QY 361 LNPEAGMMOCLLSDSGOVLLESNIKVLPTWSTPVPAPAPKSC 404  
 DB 361 LNPEAGMMOCLLSDSGOVLLESNIKVLPTWSTPVPAPAPKSC 404

ID AAR78676 standard; protein; 575 AA.  
 XX  
 XX AAR78676;  
 AC  
 XX 16-APR-1996 (first entry)  
 DT  
 XX T-cell receptor zeta.  
 DE  
 XX Chimeric receptor; CD4; T-cell receptor zeta; HIV; cytolysis;  
 KM human immunodeficiency virus; adoptive immunotherapy.  
 XX  
 XX Homo sapiens.  
 OS  
 XX MO9521528-A1.  
 PN  
 XX 17-AUG-1995.  
 PD  
 XX 12-JAN-1995; 95WO-US000454.  
 PF  
 XX 14-FEB-1994; 94US-00195395.  
 PR 02-AUG-1994; 94US-00284391.  
 XX  
 XX (GEHO ) GEN HOSPITAL CORP.  
 XX  
 XX Seed B, Banapour B, Romeo C, Kolanus W;  
 XX  
 XX WPI: 1995-292893/38.  
 XX N-PSDB; AAQ96122.  
 DR  
 XX Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing  
 PT cells.  
 PT  
 XX  
 XX Example 2; Page 76-77; 118pp; English.  
 PS  
 XX Fusion proteins comprising the extracellular domain of CD4 fused to T-  
 CC cell receptor zeta, gamma or eta (AAR78676-78, respectively) were  
 CC expressed in CV1 using a vaccine virus vector. These CD4:zeta, CD4:gamma  
 CC and CD4:eta chimeric receptors mediated cytolysis of targets expressing  
 CC HIV gp120/41  
 XX  
 XX Sequence 575 AA;  
 SQ

Query Match 59.7%; Score 2036.5; DB 2; Length 575;  
 Best Local Similarity 98.3%; Pred. No. 66-102;  
 Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;

QY 1 MNRGVPFRHLLVQLALPAATQGNKVLGKGGDTVELCTASQKSIQPHMKNNOIK 60  
 DB 1 MNRGVPFRHLLVQLALPAATQGNKVLGKGGDTVELCTASQKSIQPHMKNNOIK 60  
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120  
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120  
 QY 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
 DB 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
 QY 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
 DB 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVQCRSPRGKNIQGGKTLVSQLELDDSG 180  
 QY 181 TWCTCTVQONQKVEFKIDIVLAFOKASSIYKKEGEOVESFPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTCTVQONQKVEFKIDIVLAFOKASSIYKKEGEOVESFPLAFTVEKLTGSGELMW 240  
 QY 241 QABRASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKLPJHLTLPOALPOYAGSGNITLA 300  
 DB 241 QABRASSSSKSWITFDLKNKEVSVKRVTDPKLQMGKLPJHLTLPOALPOYAGSGNITLA 300  
 QY 301 LEATGTGLHGEVNLVVMRATQLOKNTCEVWGPTSPKLMSTLKENKEAKVSRKEKRVWV 360  
 DB 301 LEATGTGLHGEVNLVVMRATQLOKNTCEVWGPTSPKLMSTLKENKEAKVSRKEKRVWV 360  
 QY 361 LNPEAGMMOCLLSDSGOVLLESNIKVLPTWSTPVPAPAPKSC 404  
 DB 361 LNPEAGMMOCLLSDSGOVLLESNIKVLPTWSTPVPAPAPKSC 404



```
PT eliminating HIV-infected cells.
XX
XX Claim 7; Page 74-75; 120pp; English.
XX
XX A chimeric receptor (AAW00213) comprises the extracellular domain of an
CC engineered form of the CD4 cellular receptor for HIV and the
CC transmembrane and intracellular regions, including the cytosolic signal-
CC transducing portion, of the human T-cell receptor zeta chain; the region
CC of the fusion is shown in AAW02221. It can be obid. by inserting a gene
CC fusion (AAW16758) into a vaccinia virus vector and expression in host
CC cells. Chimeric receptors comprising CD4 fused to zeta, eta (see also
CC AAW02215) or Fc receptor gamma (see also AAW02214) chains are capable of
CC directing cytotoxic T lymphocytes to specifically recognise and kill
CC cells expressing HIV gp120, thus providing a therapy for AIDS
XX
XX Sequence 575 AA:
SQ
Query Match 59.7%; Score 2036.5; DB 2; Length 575;
Best Local Similarity 98.3%; Pred. No. 6e-102;
Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;
QY 1 MNRGVPRHLLLVQLALPAAIQGNKRVLGKGDVVELTCTASQKKSIOFHMKNNOIK 60
DB 1 MNRGVPRHLLLVQLALPAAIQGNKRVLGKGDVVELTCTASQKKSIOFHMKNNOIK 60
QY ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIINLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIINLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGGSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDDSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDDSG 180
QY 181 TWTCVTVLQNKQKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFVVEKLTSGGELMW 240
DB 181 TWTCVTVLQNKQKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFVVEKLTSGGELMW 240
QY 241 QAERASSSKSWITFDLNKKEVSVKRVTDQPKLQNGKPLHLTLPLQALPYAGSGNLTLLA 300
DB 241 QAERASSSKSWITFDLNKKEVSVKRVTDQPKLQNGKPLHLTLPLQALPYAGSGNLTLLA 300
QY 301 LEATGKGLHQBENLVVWRATOLQKNTLCEVWGPTSPKLMSTLLENKEAVSKREKPVW 360
DB 301 LEATGKGLHQBENLVVWRATOLQKNTLCEVWGPTSPKLMSTLLENKEAVSKREKPVW 360
QY 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPVHAA--DPKLC 401
DB 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPVHAA--DPKLC 401
RESULT 43
AAW83140
ID AAW83140 standard; protein; 575 AA.
XX
XX AAW83140;
XX
XX 03-FEB-1999 (first entry)
XX
XX Chimeric receptor containing human zeta polypeptide.
XX
XX Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;
XX tumour; cancer cell; autoimmune-generated cell; T cell receptor; CD3;
XX CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;
XX protozoan; viral.
XX
XX Synthetic.
XX OS Homo sapiens.
XX
XX US5843728-A.
XX
XX 01-DEC-1998.
XX
XX 05-APR-1995; 95US-00417495.
```

```
XX
XX 07-MAR-1991; 91US-00665961.
XX 06-MAR-1992; 92US-00847566.
XX 28-FEB-1994; 94US-00203866.
XX
XX (GENO ) GEN HOSPITAL CORP.
XX
XX Romeo C, Koltanus W, Seed B;
XX
XX WPI; 1999-044582/04.
XX N-PSDB; AAV70156.
XX
XX Membrane-bound chimeric receptors - comprising extracellular portion
XX which recognises and binds a target cell and an intracellular portion of
XX e.g. a T-cell receptor.
XX
XX Example 2; Col 39-42; 57pp; English.
XX
XX The present invention describes DNA encoding a membrane-bound chimeric
XX receptor comprising: (a) an extracellular portion that specifically
XX recognises and binds a target cell or a target infective agent; and (b)
XX an intracellular portion of a T-cell receptor CD3, zeta or eta
XX polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.
XX The present sequence represents a chimeric receptor containing the human
XX zeta polypeptide. Cells expressing chimeric receptors of the present
XX invention can be administered to mammals in order to destroy pathogens
XX (e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells
XX or autoimmune-generated cells
XX
XX Sequence 575 AA:
SQ
Query Match 59.7%; Score 2036.5; DB 2; Length 575;
Best Local Similarity 98.3%; Pred. No. 6e-102;
Matches 397; Conservative 1; Mismatches 3; Indels 3; Gaps 1;
QY 1 MNRGVPRHLLLVQLALPAAIQGNKRVLGKGDVVELTCTASQKKSIOFHMKNNOIK 60
DB 1 MNRGVPRHLLLVQLALPAAIQGNKRVLGKGDVVELTCTASQKKSIOFHMKNNOIK 60
QY ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIINLKIEDSDTYICEVEDQKEEVOL 120
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFLIINLKIEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSDTHLLQGGSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDDSG 180
DB 121 LVFGLTANSDTHLLQGGSLTLTLESPPGSSPSVQCRSPRKNIOGKTLVSQLELDDSG 180
QY 181 TWTCVTVLQNKQKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFVVEKLTSGGELMW 240
DB 181 TWTCVTVLQNKQKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFVVEKLTSGGELMW 240
QY 241 QAERASSSKSWITFDLNKKEVSVKRVTDQPKLQNGKPLHLTLPLQALPYAGSGNLTLLA 300
DB 241 QAERASSSKSWITFDLNKKEVSVKRVTDQPKLQNGKPLHLTLPLQALPYAGSGNLTLLA 300
QY 301 LEATGKGLHQBENLVVWRATOLQKNTLCEVWGPTSPKLMSTLLENKEAVSKREKPVW 360
DB 301 LEATGKGLHQBENLVVWRATOLQKNTLCEVWGPTSPKLMSTLLENKEAVSKREKPVW 360
QY 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPVHAA--DPKLC 401
DB 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPVHAA--DPKLC 401
RESULT 44
AAW27277
ID AAW27277 standard; protein; 462 AA.
XX
XX AAW27277;
XX
XX 25-MAR-2003 (revised)
XX 28-JUL-1995 (first entry)
XX
```

```

DE CD4:eta peptide chimeric protein.
XX
XX Fusion protein; CD4; extracellular domain; zeta; eta; gamma;
KM membrane spanning domain; intracellular domain; type I;
KM integral membrane homodimer; TCR; T cell antigen receptor;
KM extracellular domain; mouse; human; receptor; chimera;
KM HBB-ALL tumour cell line; natural killer cell.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Protein 1..399
FT /note="CD4 extracellular domain"
FT Protein 400..462
FT /note="zeta membrane spanning and intracellular domain"
XX
XX MO9215322-A1.
XX
XX PD 17-SEP-1992.
XX
XX PF 06-MAR-1992; 92WO-US001785.
XX
XX PR 07-MAR-1991; 91US-00665961.
XX
XX PA (GENO ) GEN HOSPITAL CORP.
XX
XX PI Seed B, Romeo C, Kolanus W;
XX
XX DR WPI; 1992-331474/40.
XX
XX DR N-PSDB; AAQ28705.
XX
XX PT Therapeutic cells expressing chimeric receptors - directing cellular
XX response to an infective agent, useful in treating HIV-1, AIDS
XX PT Pneumocystis carinii infections etc.
XX
XX Example 2; Page 73-74; 114pp; English.
XX
XX PS This sequence represents a fusion protein between the CD4 extracellular
XX domain and the eta protein membrane spanning domain and intracellular
XX domain. Eta is an isoform of zeta (see also AAR27276) which is a 32 KD
XX type I integral membrane homodimer, which arises by alternate RNA
XX splicing. It is present in reduced amounts in cells expressing the T cell
XX antigen receptor. Zeta-eta heterodimers are thought to mediate the
XX formation of inositol phosphates, as well as the receptor initiated cell
XX death called apoptosis. In the production of the CD4 receptor initiated cell
XX the eta cDNA was isolated from the HBB-ALL tumour cell line and from
XX human natural killer cells. The eta cDNA was joined to the extracellular
XX domain of an engineered form of CD4 possessing a BamHI site just upstream
XX of the membrane spanning domain, by a BamHI site naturally present a few
XX residues upstream of the membrane spanning domain. (Updated on 25-MAR-
XX 2003 to correct PN field.)
XX
XX SQ Sequence 462 AA;
XX
XX Query Match 59.5%; Score 2032.5; DB 2; Length 462;
XX Best Local Similarity 98.0%; Pred. No. 7.8e-102;
XX Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;
XX
XX QY 1 MNRGVFRRHLVLTQALPAATQGNKRVLGKGGDTVELTCTASOKKSIQFHKNSNQIK 60
XX DB 1 MNRGVFRRHLVLTQALPAATQGNKRVLGKGGDTVELTCTASOKKSIQFHKNSNQIK 60
XX
XX QY 61 ILNGQSFLLTKGPSKLNDRADRSRSLMDQGNFPLIINKLKI ESDTYICEVEDQKEEVOL 120
XX DB 61 ILNGQSFLLTKGPSKLNDRADRSRSLMDQGNFPLIINKLKI ESDTYICEVEDQKEEVOL 120
XX
XX QY 121 LVFGLTANSDTHLQOGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLISVSQL ELDQSG 180
XX DB 121 LVFGLTANSDTHLQOGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLISVSQL ELDQSG 180
XX
XX QY 181 TWFTCTVQONQKVEFKDIIIVLAFQKASSIYKKKGQVEFFSPFLATVEKLTGSGELMW 240
XX DB 181 TWFTCTVQONQKVEFKDIIIVLAFQKASSIYKKKGQVEFFSPFLATVEKLTGSGELMW 240

```

```

QY 241 QAEARSSSKSWITTEPDLKXKVEVSRVTDPPKLOMKKXLPILHTLPQALPOVAGSGNLTLA 300
DB 241 QAEARSSSKSWITTEPDLKXKVEVSRVTDPPKLOMKKXLPILHTLPQALPOVAGSGNLTLA 300
QY 301 LEAKTGKTLHQBENLVVMBATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVVW 360
DB 301 LEAKTGKTLHQBENLVVMBATQLOKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVVW 360
QY 361 LNPEAGMWOCLISDSGOVLBSNIVLPTWSTPVPCEPAEPKSC 404
DB 361 LNPEAGMWOCLISDSGOVLBSNIVLPTWSTPVPCEPAEPKSC 404
XX
XX RESULT 45
XX ID AAR78677 standard; protein; 462 AA.
XX
XX AC AAR78677;
XX
XX DT 16-APR-1996 (first entry)
XX
XX DE T-cell receptor gamma.
XX
XX KM Chimeric receptor; CD4; T-cell receptor gamma; HIV; cytolysis;
XX human immunodeficiency virus; adoptive immunotherapy.
XX
XX OS Homo sapiens.
XX
XX OS MO9521528-A1.
XX
XX PD 17-AUG-1995.
XX
XX PF 12-JAN-1995; 95WO-US000454.
XX
XX PR 14-FEB-1994; 94US-00195395.
XX
XX PR 02-AUG-1994; 94US-00284391.
XX
XX PA (GENO ) GEN HOSPITAL CORP.
XX
XX PI Seed B, Banapour B, Romeo C, Kolanus W;
XX
XX DR WPI; 1995-292893/38.
XX
XX DR P-PSDB; AAQ96123.
XX
XX PT Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing
XX cells.
XX
XX PS Example 2; Page 77-78; 118pp; English.
XX
XX CC Fusion proteins comprising the extracellular domain of CD4 fused to T-
XX cell receptor zeta, gamma or eta (AAR78676-78, respectively) were
XX expressed in CV1 using a vaccine virus vector. These CD4:zeta, CD4:gamma
XX and CD4:eta chimeric receptors mediated cytolysis of targets expressing
XX HIV gp120/41
XX
XX SQ Sequence 462 AA;
XX
XX Query Match 59.5%; Score 2032.5; DB 2; Length 462;
XX Best Local Similarity 98.0%; Pred. No. 7.8e-102;
XX Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;
XX
XX QY 1 MNRGVFRRHLVLTQALPAATQGNKRVLGKGGDTVELTCTASOKKSIQFHKNSNQIK 60
XX DB 1 MNRGVFRRHLVLTQALPAATQGNKRVLGKGGDTVELTCTASOKKSIQFHKNSNQIK 60
XX
XX QY 61 ILNGQSFLLTKGPSKLNDRADRSRSLMDQGNFPLIINKLKI ESDTYICEVEDQKEEVOL 120
XX DB 61 ILNGQSFLLTKGPSKLNDRADRSRSLMDQGNFPLIINKLKI ESDTYICEVEDQKEEVOL 120
XX
XX QY 121 LVFGLTANSDTHLQOGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLISVSQL ELDQSG 180
XX DB 121 LVFGLTANSDTHLQOGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLISVSQL ELDQSG 180

```



```

QY 181 TWCTVLONOKKVEFKIDIVLAFQKASSIYKKKEGQVEFSFPLAFTVEKLTGSGELMW 240
    |||||
Db 181 TWCTVLONOKKVEFKIDIVLAFQKASSIYKKKEGQVEFSFPLAFTVEKLTGSGELMW 240
QY 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQWKKLPLHLTLPQALPOYAGSGNLTLA 300
    |||||
Db 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQWKKLPLHLTLPQALPOYAGSGNLTLA 300
QY 301 LEATGTLHQBVLVWVRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKPVWV 360
    |||||
Db 301 LEATGTLHQBVLVWVRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKPVWV 360
QY 361 LNPAGMWOCLLSDSGOVLLESNIKVLPTWSTPVPCEAPBPKSC 404
    |||||
Db 361 LNPAGMWOCLLSDSGOVLLESNIKVLPTWSTPVPCEAPBPKSC 404

RESULT 46
AAR89457
ID AAR89457 standard; protein; 462 AA.
XX
AC AAR89457;
XX
DT 26-SEP-1996 (first entry)
XX
DE CD4:gamma fusion protein.
XX
KM CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;
    human; CD4; HIV; proteinaceous alpha-helix; T cell; B cell; neutrophil;
    dendritic cell; therapy; mammal; infection.
XX
OS Synthetic.
XX
PN WO9603883-A1.
XX
PD 15-FEB-1996.
XX
PF 26-JUL-1995; 95WO-US009468.
XX
PR 02-AUG-1994; 94US-00284391.
    24-FEB-1995; 95US-00394388.
XX
PA (GENO) GEN HOSPITAL CORP.
XX
PI Seed B, Banapour B, Romeo C, Kolanus W;
XX
DR WPI; 1996-129034/13.
    N-PSDB; AAT10802.
XX
PT Membrane-bound chimeric receptor comprising extracellular portion
    including CD4 fragment - cells expressing receptor can be used for
    treatment of HIV infection.
XX
PS Example 2; Page 79; 134pp; English.
XX
CC AAT10801-T10803 represent membrane bound proteinaceous chimeric receptors
    of the invention. This sequence represents the CD4:gamma chimera. The
    transmembrane region of the chimeric receptor acts to separate the
    intracellular and extracellular domains of the chimera, and contains a
    portion of the CD7 (see AAR89440), CD5 or CD34 transmembrane domains.
    Alternatively, the extracellular portion of the receptor can be separated
    from the intracellular domain by the hinge, CH2 and CH3 domains of human
    IgG1 (see AAR89441). The extracellular portion of the chimeric receptor
    contains a fragment of CD4 (amino acids 1-394 or 1-200 of the CD4
    sequence, see AAR89450 and AAR89451) which specifically recognizes and
    binds HIV-infected cells, but does not mediate HIV infection. The
    extracellular domain of the receptor is separated from the cell membrane
    by 48 or 72 angstroms, or by one or more proteinaceous alpha-helices.
    The cells expressing the receptor are preferably T cells, B cells,
    neutrophils, or dendritic cells. The therapeutic cells expressing the
    chimeric receptor are administered to a mammal to treat HIV infection
  
```

```

SQ Sequence 462 AA:
Query Match 59.5%; Score 2032.5; DB 2; Length 462;
Best Local Similarity 98.0%; Pred. No. 7.8e-102;
Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;

QY 1 MNRGVPRHLHLVQLALPAAATQGNKVYLGKGDVETLCTASQKKSIOFHMKNNOIK 60
    |||||
Db 1 MNRGVPRHLHLVQLALPAAATQGNKVYLGKGDVETLCTASQKKSIOFHMKNNOIK 60
QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFLLIKNLIKEDSDTYICEVEDQKEEVOL 120
    |||||
Db 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFLLIKNLIKEDSDTYICEVEDQKEEVOL 120
QY 121 LVFGLTANSPDTHLLQGGSLTLTLSPGSSPSVQCRRPRKNTIQGKTLISVSQLELDGSG 180
    |||||
Db 121 LVFGLTANSPDTHLLQGGSLTLTLSPGSSPSVQCRRPRKNTIQGKTLISVSQLELDGSG 180
QY 181 TWCTVLONOKKVEFKIDIVLAFQKASSIYKKKEGQVEFSFPLAFTVEKLTGSGELMW 240
    |||||
Db 181 TWCTVLONOKKVEFKIDIVLAFQKASSIYKKKEGQVEFSFPLAFTVEKLTGSGELMW 240
QY 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQWKKLPLHLTLPQALPOYAGSGNLTLA 300
    |||||
Db 241 QAERASSSSKSWITFDLKNKEVSVKRVTDPKLQWKKLPLHLTLPQALPOYAGSGNLTLA 300
QY 301 LEATGTLHQBVLVWVRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKPVWV 360
    |||||
Db 301 LEATGTLHQBVLVWVRATQLOKNTLCEVWGPTSPKLMSTLKENKEAKVSKREKPVWV 360
QY 361 LNPAGMWOCLLSDSGOVLLESNIKVLPTWSTPVPCEAPBPKSC 404
    |||||
Db 361 LNPAGMWOCLLSDSGOVLLESNIKVLPTWSTPVPCEAPBPKSC 404

RESULT 47
AAM02214
ID AAM02214 standard; protein; 462 AA.
XX
AC AAM02214;
XX
DT 11-NOV-1996 (first entry)
XX
DE CD4:Fc receptor gamma chain chimaeric receptor.
XX
KM Chimaeric receptor; cellular immunity; adoptive immunotherapy; CD4;
    human immunodeficiency virus type 1; HIV-1; AIDS; therapy;
    Fc receptor gamma chain; cytotoxic T lymphocyte; CTL.
XX
OS Homo sapiens.
XX
FH Key
    Location/Qualifiers
    1..393
    /label= "Extracellular domain"
    /note= "CD4 extracellular domain"
    394..397
    /label= "linker"
    /note= "encoding DNA contains a BamHI site used for
    fusion construction"
    398..462
    /note= "region of fusion derived from gamma chain,
    preferred signal-transducing portions for constructs of
    the invention are amino acids 421-462 and 402-419"
    400..462
    /label= "Transmembrane+intracellular domains"
XX
PN WO9625953-A1.
XX
PD 29-AUG-1996.
XX
PR 25-JAN-1996; 96WO-US001056.
    24-FEB-1995; 95US-00394176.
  
```

XX (GEHO ) GEN HOSPITAL CORP.  
 PA Seed B, Romeo C, Kolanus W;  
 XX WPI; 1996-402134/40.  
 DR N-PSDB; AAT36759.  
 DR  
 XX Direction of cellular immune response using therapeutic cell expressing 2  
 PT chimeric receptors - comprising region binding to target cell and region  
 PT that signals target cell destruction, or CD28 region, partic. for  
 PT eliminating HIV-infected cells.  
 XX  
 PS Claim 7, Page 76; 120pp; English.  
 XX  
 XX A chimeric receptor (AAW0214) comprises the extracellular domain of an  
 CC engineered form of the CD4 cellular receptor for HIV and the  
 CC transmembrane and intracellular regions, including the cytosolic signal-  
 CC transducing portion, of the human Fc receptor gamma chain; the region of  
 CC the fusion is shown in AAW0223. It can be obt'd. by inserting a gene  
 CC fusion (AAT36759) into a vaccinia virus vector and expressing in host  
 CC cells. Chimeric receptors comprising CD4 fused to Fc receptor gamma or T  
 CC -cell receptor zeta (see also AAW02213) or eta (AAW02215) chains are  
 CC capable of directing cytotoxic T lymphocytes to specifically recognise  
 CC and kill cells expressing HIV gp120, thus providing a therapy for AIDS  
 CC  
 XX Sequence 462 AA;  
 SQ  
 Query Match 59.5%; Score 2032.5; DB 2; Length 462;  
 Best Local Similarity 98.0%; Pred. No. 7.8e-102;  
 Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;  
 QY 1 MNRGVPRHLLVLTQALLPAATQGNKVVIGKGGDTVELTCTASQKKSIOFHWKNSNQIK 60  
 DB 1 MNRGVPRHLLVLTQALLPAATQGNKVVIGKGGDTVELTCTASQKKSIOFHWKNSNQIK 60  
 QY 61 ILGNQGSFLTTPSPSKLNDADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120  
 DB 61 ILGNQGSFLTTPSPSKLNDADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120  
 QY 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSSVQLBLQDSG 180  
 DB 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSSVQLBLQDSG 180  
 QY 181 TWCTCTVLQNOQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTCTVLQNOQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPPLAFTVEKLTGSGELMW 240  
 QY 241 QABRASSKSWITFDLKNKEVSVKRYTOPPKLOMGKKLPLHLTLPLPALPOYAGSGNLTLLA 300  
 DB 241 QABRASSKSWITFDLKNKEVSVKRYTOPPKLOMGKKLPLHLTLPLPALPOYAGSGNLTLLA 300  
 QY 301 LEAKTGKLEHENVLVVWRATOLQKNLTCCEWGPSPKMLSLKLENKEAKVSREKRPVWV 360  
 DB 301 LEAKTGKLEHENVLVVWRATOLQKNLTCCEWGPSPKMLSLKLENKEAKVSREKRPVWV 360  
 QY 361 LNPBAGMWQCLLSDSGVLLSNIKVLPTWSTPVPAPBPXSC 404  
 DB 361 LNPBAGMWQCLLSDSGVLLSNIKVLPTWSTPVPAPBPXSC 404  
 Db 361 LNPBAGMWQCLLSDSGVLLSNIKVLPTWSTPVPAPBPXSC 404

## RESULT 48

AAW83142 standard; protein; 462 AA.

AAW83142;

XX 03-FEB-1999 (first entry)  
 XX Chimeric receptor containing mouse gamma polypeptide.  
 DE Human; zeta; eta; gamma; membrane-bound chimeric receptor; infection;  
 KM tumour; cancer cell; autoimmune-generated cell; T cell receptor; CD3;

KM CD4; B cell receptor; Fc receptor; pathogen; bacterial; fungal;  
 KM protozoan; viral.  
 XX Synthetic.  
 OS Mus sp.  
 OS  
 XX US5843728-A.  
 XX  
 PD 01-DEC-1998.  
 XX  
 PF 05-APR-1995; 95US-00417495.  
 XX  
 PR 07-MAR-1991; 91US-00665961.  
 PR 06-MAR-1992; 92US-00847566.  
 PR 28-FEB-1994; 94US-00203866.  
 XX  
 XX (GEHO ) GEN HOSPITAL CORP.  
 XX  
 PI Romeo C, Kolanus W, Seed B;  
 XX WPI; 1999-044582/04.  
 DR N-PSDB; AAV70158.  
 XX  
 PT Membrane-bound chimeric receptors - comprising extracellular portion  
 PT which recognises and binds a target cell and an intracellular portion of  
 PT e.g. a T-cell receptor.  
 XX  
 PS Example 2; Col 43-46; 57pp; English.  
 XX  
 XX The present invention describes DNA encoding a membrane-bound chimeric  
 CC receptor comprising: (a) an extracellular portion that specifically  
 CC recognises and binds a target cell or a target infective agent; and (b)  
 CC an intracellular portion of a T-cell receptor CD3, zeta or eta  
 CC polypeptide, a B-cell receptor polypeptide or an Fc receptor polypeptide.  
 CC The present sequence represents a chimeric receptor containing the mouse  
 CC gamma polypeptide. Cells expressing chimeric receptors of the present  
 CC invention can be administered to mammals in order to destroy pathogens  
 CC (e.g. bacteria, fungi, protozoa or viruses, especially HIV), cancer cells  
 CC or autoimmune-generated cells  
 CC  
 XX Sequence 462 AA;  
 SQ  
 Query Match 59.5%; Score 2032.5; DB 2; Length 462;  
 Best Local Similarity 98.0%; Pred. No. 7.8e-102;  
 Matches 396; Conservative 2; Mismatches 3; Indels 3; Gaps 1;  
 QY 1 MNRGVPRHLLVLTQALLPAATQGNKVVIGKGGDTVELTCTASQKKSIOFHWKNSNQIK 60  
 DB 1 MNRGVPRHLLVLTQALLPAATQGNKVVIGKGGDTVELTCTASQKKSIOFHWKNSNQIK 60  
 QY 61 ILGNQGSFLTTPSPSKLNDADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120  
 DB 61 ILGNQGSFLTTPSPSKLNDADSRSLMDQGNFPLIINKLIEDSDTYICEVEDQKEVOL 120  
 QY 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSSVQLBLQDSG 180  
 DB 121 LVFGLTANSDTHLLOQOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLSSVQLBLQDSG 180  
 QY 181 TWCTCTVLQNOQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTCTVLQNOQKVEFKIDIVLAFOKASSIVYKKEGQVEFSPPPLAFTVEKLTGSGELMW 240  
 QY 241 QABRASSKSWITFDLKNKEVSVKRYTOPPKLOMGKKLPLHLTLPLPALPOYAGSGNLTLLA 300  
 DB 241 QABRASSKSWITFDLKNKEVSVKRYTOPPKLOMGKKLPLHLTLPLPALPOYAGSGNLTLLA 300  
 QY 301 LEAKTGKLEHENVLVVWRATOLQKNLTCCEWGPSPKMLSLKLENKEAKVSREKRPVWV 360  
 DB 301 LEAKTGKLEHENVLVVWRATOLQKNLTCCEWGPSPKMLSLKLENKEAKVSREKRPVWV 360  
 QY 361 LNPBAGMWQCLLSDSGVLLSNIKVLPTWSTPVPAPBPXSC 404  
 DB 361 LNPBAGMWQCLLSDSGVLLSNIKVLPTWSTPVPAPBPXSC 404

ID	AAW02215	standard; protein; 532 AA.
AC	AAW02215;	
DT	16-OCT-2003	(revised)
DT	11-NOV-1996	(first entry)
DE	CD4-T-cell receptor eta chain chimeric receptor.	
XX	Chimeric receptor; cellular immunity; adoptive immunotherapy; CD4;	
KM	human immunodeficiency virus type 1; HIV-1; AIDS; therapy;	
KW	T-cell receptor eta chain; cytotoxic T lymphocyte; CTL.	
OS	Homo; sapiens.	
OS	Mus sp.	
OS	Chimeric.	
PH	Key	Location/Qualifiers
FT	Domain	1..393
FT		/label="Extracellular domain
FT		/note="CD4 extracellular domain"
FT	Region	394..396
FT		/label="Linker
FT		/note="encoding DNA contains a BamHI site used for fusion construction"
FT	Region	397..532
FT		/note="region of fusion derived from eta chain, preferred signal-transducing portions for constructs of the invention are amino acids 421-532, 423-455, 438-455, 461-494, 494-528 or 400-420"
FT	Domain	400..437
FT		/label="Transmembrane domain
FT		/note="eta chain transmembrane domain"
FT	Domain	438..575
FT		/label="Intracellular domain
FT		/note="eta chain intracellular domain"
PN	MO9625953-A1.	
PD	29-AUG-1996.	
PF	25-JAN-1996;	96WO-US001056.
PR	24-FEB-1995;	95US-00394176.
PA	(GENO ) GEN HOSPITAL CORP.	
PI	Seed B, Romeo C, Kolanus W;	
DR	WPI; 1996-402134/40.	
DR	N-PSDB; AAT36760.	
PT	Direction of cellular immune response using therapeutic cell expressing 2	
PT	chimeric receptors - distinguishing region binding to target cell and region	
PT	that signals target cell destruction, or CD28 region, partic. for	
PT	eliminating HIV-infected cells.	
PS	Claim 7; Page 77-78; 120pp; English.	
CC	A chimeric receptor (AAW00215) comprises the extracellular domain of an engineered form of the CD4 cellular receptor for HIV and the transmembrane and intracellular regions, including the cytoytic signal-transducing portion, of the mouse T-cell receptor eta chain. It can be obtd. by inserting a gene fusion (AAT36760) into a vaccinia virus vector and expression in host cells. Chimeric receptors comprising CD4 fused to eta, eta (see also AAW02213) or Fc receptor gamma (see also AAW02214) chains are capable of directing cytotoxic T lymphocytes to specifically recognise and kill cells expressing HIV gp120, thus providing a therapy for AIDS. (Updated on 16-Oct-2003 to standardise OS field)	

XX	Sequence	532 AA:	59.5%;	Score 2032.5;	DB 2;	Length 532;
XX	Query Match		98.0%;	Pred. No. 9.1e-102;		
XX	Best Local Similarity		98.0%;	Pred. No. 9.1e-102;		
XX	Matches 396;	Conservative 2;	Mismatches 3;	Indels 3;	Gaps 1;	
QY	1	MNRGVPPRHLLLVQLALLPAATQGNRVVLGKKDVTVELTCTASQKKSIOFHMKNSNOIK	60			
DB	1	MNRGVPPRHLLLVQLALLPAATQGNRVVLGKKDVTVELTCTASQKKSIOFHMKNSNOIK	60			
QY	61	ILGNQGSFLTYGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEYQL	120			
DB	61	ILGNQGSFLTYGSPSKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEEYQL	120			
QY	121	LVFGLTANSDDTHLLQGGSLTTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG	180			
DB	121	LVFGLTANSDDTHLLQGGSLTTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSG	180			
QY	181	TWCTCTVLOQNKKEVFKDIDVVLAAQKASSIYYKKEGSEVSPFLATVEKLTSSGELMW	240			
DB	181	TWCTCTVLOQNKKEVFKDIDVVLAAQKASSIYYKKEGSEVSPFLATVEKLTSSGELMW	240			
QY	241	QAEKASSSKSWITFDLANKKEVSVKRVYQDPKLOMGKTLPLHLTLPOALPOYAGSGNLTIA	300			
DB	241	QAEKASSSKSWITFDLANKKEVSVKRVYQDPKLOMGKTLPLHLTLPOALPOYAGSGNLTIA	300			
QY	301	LEAKTGKLEHENVLVNWRATOLQKNTLCEVWGPTSPKLMSTLKENKBAKVSKEKPYMV	360			
DB	301	LEAKTGKLEHENVLVNWRATOLQKNTLCEVWGPTSPKLMSTLKENKBAKVSKEKPYMV	360			
QY	361	LNPEAGMWQCLSDSGOVLLESNIKVLPTWSTPLVPCPAPEPKXC	404			
DB	361	LNPEAGMWQCLSDSGOVLLESNIKVLPTWSTPLVPCPAPEPKXC	404			
XX	RESULT 50					
XX	AAR89450					
XX	AAR89450	standard; peptide; 398 AA.				
XX	AC	AAR89450;				
XX	DT	26-SEP-1996 (first entry)				
XX	DE	CD4 DI-D4 domains.				
XX	KW	CD7; transmembrane domain; chimeric receptor; CD5; CD34; CH2; CH3; IgG1;				
XX	KW	human; CD4; HIV; proteolaceous alpha-helix; T cell; B cell; neutrophil;				
XX	OS	dendritic cell; therapy; mammal; infection.				
XX	PN	Homo sapiens.				
XX	PD	WO9603883-A1.				
XX	PF	26-JUL-1995; 95WO-US009468.				
XX	PR	02-AUG-1994; 94US-00284391.				
XX	PA	24-FEB-1995; 95US-00394388.				
XX	PI	(GEHO ) GEN HOSPITAL CORP.				
XX	PI	Seed B, Banapour B, Romeo C, Kolanus W;				
XX	DR	WPI; 1996-129034/13.				
XX	DR	N-PSDB; AAT10797.				
XX	PT	Membrane-bound chimeric receptor comprising extracellular portion				
XX	PT	including CD4 fragment - cells expressing receptor can be used for				
XX	PS	treatment of HIV infection.				
XX	PS	Example 10; Fig 23; 134dp; English				



Qy 361 LNPEAGMWCCLSDSGVLLSNIKVLPTWSTPV 394  
 Db 361 LNPEAGMWCCLSDSGVLLSNIKVLPTWSTPV 394

## RESULT 52

AAR78673  
 ID AAR78673 standard; protein; 398 AA.

XX AAR78673;

XX 12-APR-1996 (first entry)

XX CD4 domains D1-D4.

XX Chimeric receptor; CD4; T-cell receptor; HIV; cytolysis;

XX human immunodeficiency virus; adoptive immunotherapy.

XX Homo sapiens.

XX MO9521528-A1.

XX 17-AUG-1995.

XX 12-JAN-1995; 95WO-US000454.

XX 14-FEB-1994; 94US-00195395.

XX 02-AUG-1994; 94US-00284391.

XX (GCHO ) GEN HOSPITAL CORP.

XX Seed B, Banapur B, Romeo C, Kolanus W;

XX WPI; 1995-292893/38.

XX N-PSDB; AAQ96103.

XX Target cytolysis of HIV-infected cells - by chimeric CD4 receptor-bearing cells.

XX Example 10; Fig 23; 118pp; English.

XX Extracellular domains D1-D4 (AAR78673) or D1-D2 (AAR78674) of human CD4

XX are used in the construction of chimeric receptors utilised in the

XX targeted cytolysis of cells expressing HIV envelope proteins on their

XX surface. The chimeric receptors comprise the extracellular domain (pref.

XX amino acids 1-394 or 1-200) of CD4 linked to an intracellular portion,

XX e.g. of T-cell receptor protein zeta

XX Sequence 398 AA;

XX Query Match 59.3%; Score 2026; DB 2; Length 398;

XX Best Local Similarity 99.7%; Pred. No. 1.5e-101;

XX Matches 393; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MNRGVPFHLVLLVQALLPAATQGNKVVLGKKGDTVELTCTASOKSIQFHMKNINQIX 60

Db 1 MNRGVPFHLVLLVQALLPAATQGNKVVLGKKGDTVELTCTASOKSIQFHMKNINQIX 60

Qy 61 ILGNQGSFLTKGPSKLRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEYQL 120

Db 61 ILGNQGSFLTKGPSKLRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEYQL 120

Qy 121 LVFGITANSDFHLLOGOSLTLTLESPPSSPSVQCRSRGNKIQGGKTLVSQLELQDSG 180

Db 121 LVFGITANSDFHLLOGOSLTLTLESPPSSPSVQCRSRGNKIQGGKTLVSQLELQDSG 180

Qy 181 TMTCTVLNOKKVEFKIDIVLAFKASSIYKKKEGEVSPFLAFTVEKLTGSGELMW 240

Db 181 TMTCTVLNOKKVEFKIDIVLAFKASSIYKKKEGEVSPFLAFTVEKLTGSGELMW 240

Qy 301 LEAKTKLHQBENVLVNRATOLQKLTCEVGPPTSPLMLSLKLENBAKYSKREKPYMV 360  
 Db 301 LEAKTKLHQBENVLVNRATOLQKLTCEVGPPTSPLMLSLKLENBAKYSKREKPYMV 360

## RESULT 53

AAB19509  
 ID AAB19509 standard; protein; 416 AA.

XX AAB19509;

XX 09-JAN-2001 (first entry)

XX CD4-IGM fusion protein CH4Mmu.

XX CD4; IGM; human; CD4Mmu; fusion protein; immunoglobulin; HIV; SIV; gp120;

XX therapy; diagnosis.

XX Homo sapiens.

XX Key Location/Qualifiers

XX Protein 1..395

XX FT /note= "CD4 extracellular region"

XX FT Protein 400..416

XX FT /note= "IGM heavy chain partial sequence"

XX US6117656-A.

XX 12-SEP-2000.

XX 07-JUN-1995; 95US-00479353.

XX 22-JAN-1988; 88US-00147351.

XX 23-JAN-1989; 89US-00299596.

XX 09-JUN-1992; 92US-00896781.

XX 12-APR-1993; 93US-00057952.

XX 04-FEB-1994; 94US-00191708.

XX (GCHO ) GEN HOSPITAL CORP.

XX Seed B;

XX WPI; 2000-586558/55.

XX N-PSDB; AAA50662.

XX CD4-immunoglobulin fusion proteins, useful for targeting gp120 of HIV or

XX SIV.

XX Example 1; Col 41-50; 39pp; English.

XX The present sequence is that of fusion protein CD4Mmu comprising the

XX extracellular portion of CD4, which binds to HIV gp120, linked at its C-

XX terminus to the human IGM heavy chain. To obtain the fusion protein, DNA

XX encoding CD4 was linked to IGM DNA at the MscI site upstream of the CH1

XX region (see AAA50662). Fusion protein CD4Mmu and a nucleic acid encoding

XX it are claimed. Also claimed are a vector comprising the nucleic acid,

XX and a method of producing the fusion protein in secreted form using a

XX transformed host cell. The fusion protein may further comprise a

XX therapeutic agent, radiolabel or NMR imaging agent. The fusion protein

XX can be administered to an animal (including humans) for treatment of HIV

XX or SIV infection, and can also be used in assays for HIV or SIV, imaging

XX and tissue stains. IGM fusion proteins such as CD4Mmu provide complement-

XX mediated immunity

XX Sequence 416 AA;

XX Query Match 59.3%; Score 2026; DB 3; Length 416;

XX Best Local Similarity 94.6%; Pred. No. 1.6e-101;

Matches 401; Conservative 0; Mismatches 7; Indels 16; Gaps 3;

QY 1 MNRGVPRRLILVLOALIPATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60  
 DB 1 MNRGVPRRLILVLOALIPATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60  
 QY 61 ILGNQSPFLTKGPKSKLNDRADSRSLMDQGNFPLIIKNKIEDSDTYICEVEDQKEEVL 120  
 DB 61 ILGNQSPFLTKGPKSKLNDRADSRSLMDQGNFPLIIKNKIEDSDTYICEVEDQKEEVL 120  
 QY 121 LVFGLTANSPTHLQOGSLTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSPTHLQOGSLTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWTCVLYQNKQKVEPKIDIVLAFOKASSIVYKKEGQVEFSPLAFYVEKLTGSGELMW 240  
 DB 181 TWTCVLYQNKQKVEPKIDIVLAFOKASSIVYKKEGQVEFSPLAFYVEKLTGSGELMW 240  
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKLLPLHLTLPOLPOYAGSGNLTLLA 300  
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKLLPLHLTLPOLPOYAGSGNLTLLA 300  
 QY 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVVY 360  
 DB 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVVY 360  
 QY 361 LNPBAGMOCCLSDSGOVLLESNIKVLPTWSTPVPKAPKPCDKTHTEPELLGSPV 419  
 DB 361 LNPBAGMOCCLSDSGOVLLESNIKVLPTWSTPVPKAPKPCDKTHTEPELLGSPV 419  
 QY 420 FLFP 423  
 DB 406 TLFP 409

RESULT 54  
 AAR6374  
 ID AAR6374 standard; protein; 400 AA.  
 XX  
 AC AAR6374;  
 XX  
 DT 31-OCT-2002 (revised)  
 DT 20-DEC-1990 (first entry)  
 XX  
 DE Truncated form of soluble T4 encoded by pBc381.  
 XX  
 KM plasmid pBc381; soluble T4 protein; AIDS; ARC; HIV.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..23  
 FT /label= "secretory signal"  
 FT /note= "hydrophobic"  
 FT 24..117  
 FT Region /label= "extracellular"  
 FT /note= "homology to V-regions"  
 FT 118..132  
 FT Region /label= "extracellular"  
 FT /note= "homology to J-regions"  
 FT 133..397  
 FT Region /label= "extracellular"  
 FT /note= "glycosylated"  
 XX  
 XX WO9008198-A.  
 XX  
 XX PD 26-JUL-1990.  
 XX  
 XX PF 18-JAN-1989; 89US-00300096.  
 XX  
 XX PR 18-JAN-1989; 89US-00300096.  
 XX  
 PA (HARD ) HARVARD COLLEGE.

XX  
 PI Letvin NA;  
 XX  
 DR WPI; 1990-254040/33.  
 DR N-PSDB; AAO05608.  
 XX  
 PT Treating or preventing AIDS, ARC or HIV infection - by administering an  
 PT immunologically effective amt. of soluble T4 protein.  
 XX  
 PS Disclosure; Fig 2; 121pp; English.  
 XX  
 CC T4-encoding plasmid pBc381 was used to transform Chinese Hamster Ovary  
 CC cells for the production of truncated T4. Soluble T4 is produced by  
 CC virtue of the removal of the transmembrane and cytoplasmic domains. The  
 CC soluble forms may be modified to increase their immunogenicity by  
 CC addition of an adjuvant such as incomplete Freund's adjuvant. The T4  
 CC interferes with HIV/T4 interaction and elicits anti-soluble T4 antibody  
 CC production. See also AAO05607. (Updated on 31-OCT-2002 to add missing OS  
 CC field.)  
 XX  
 SO Sequence 400 AA;

Query Match 59.2%; Score 2021; DB 2; Length 400;  
 Best Local Similarity 99.7%; Pred. No. 2.8e-101;  
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVPRRLILVLOALIPATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60  
 DB 1 MNRGVPRRLILVLOALIPATQGNKVLGKKGDVVELTCTASQKKSIOFHMKNSNQIK 60  
 QY 61 ILGNQSPFLTKGPKSKLNDRADSRSLMDQGNFPLIIKNKIEDSDTYICEVEDQKEEVL 120  
 DB 61 ILGNQSPFLTKGPKSKLNDRADSRSLMDQGNFPLIIKNKIEDSDTYICEVEDQKEEVL 120  
 QY 121 LVFGLTANSPTHLQOGSLTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSPTHLQOGSLTLTLESPGSSPVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWTCVLYQNKQKVEPKIDIVLAFOKASSIVYKKEGQVEFSPLAFYVEKLTGSGELMW 240  
 DB 181 TWTCVLYQNKQKVEPKIDIVLAFOKASSIVYKKEGQVEFSPLAFYVEKLTGSGELMW 240  
 QY 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKLLPLHLTLPOLPOYAGSGNLTLLA 300  
 DB 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKLQMGKLLPLHLTLPOLPOYAGSGNLTLLA 300  
 QY 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVVY 360  
 DB 301 LEAKTGKLEHVEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAVSKREKPVVY 360  
 QY 361 LNPBAGMOCCLSDSGOVLLESNIKVLPTWSTPVP 394  
 DB 361 LNPBAGMOCCLSDSGOVLLESNIKVLPTWSTPVP 394

RESULT 55  
 AAP81990  
 ID AAP81990 standard; protein; 458 AA.  
 XX  
 AC AAP81990;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 15-OCT-1990 (first entry)  
 XX  
 DE Clone pT4B encoded HIV T4 glycoprotein.  
 XX  
 XX KM Human Immunodeficiency Virus T4 envelope glycoprotein; AIDS; ss.  
 XX  
 OS Synthetic.  
 XX  
 FH Key Location/Qualifiers  
 FT Peptide 1..23  
 FT /label= leader peptide

FT	Region	24..117
FT	/label= variable-like region 1	
FT	Region	118..134
FT	/label= joining-like region 1	
FT	Region	135..189
FT	/label= variable-like region 2	
FT	Region	190..205
FT	/label= joining-like region 2	
FT	Region	206..287
FT	/label= variable-like region 3	
FT	Region	288..309
FT	/label= joining-like region 3	
FT	Modified-site	297..299
FT	/label= putative N-glycosylation site	
FT	Region	310..377
FT	/label= variable-like region 4	
FT	Modified-site	325..327
FT	/label= putative N-glycosylation site	
FT	Region	378..397
FT	/label= joining-like region 4	
FT	Region	398..420
FT	/label= trans-membrane region	
FT	Region	421..458
FT	/label= cytoplasmic region	
XX		
XX	WC8801304-A.	
PN		
PD	25-FEB-1988.	
XX		
PF	20-AUG-1987;	87WO-US002050.
XX		
PR	21-AUG-1986;	86US-00898587.
XX		
PA	(UYCO-) COLUMBIA UNIV.	
PA	(MADD/) MADDON P J.	
PI	Littman DR, Maddon PJ, Chess L, Axel R, Weiss R, McDougal JS;	
DR	WPI, 1988-064019/09.	
DR	N-PSDB; AAN80512.	
XX		
PT	Nucleic acid encoding T4 glyco:protein - used for treatment of AIDS and	
PT	producing antibodies for use as vaccine for immunisation against AIDS.	
XX		
PS	Disclosure; Page 7, 128pp; English.	
XX		
CC	T4 protein encoded by part of 3kb insert from human T cell library	
CC	(pT4B). (Updated on 25-MAR-2003 to correct PA field.)	
XX		
IQ	Sequence 458 AA;	

Query Match	59.2%	Score 2021	DB 1	Length 458
Best Local Similarity	99.7%	Pred. No. 3.2e-101		
Matches 393	Conservative	0	Mismatches 1	Indels 0
Gaps				0
QY	1	MNRGPPFPHLLLVLTALLPAAQGNKVLGGKGDVETLTCAQSKSIQFHMKNSNIK	60	
Db	1	MNRGPPFPHLLLVLTALLPAAQGNKVLGGKGDVETLTCAQSKSIQFHMKNSNIK	60	
QY	61	ILGNQGSFLTGPSTKLNDRADRSRLWQGNFPLIIKNLKIEDSTTYICEVEDQKEVOL	120	
Db	61	ILGNQGSFLTGPSTKLNDRADRSRLWQGNFPLIIKNLKIEDSTTYICEVEDQKEVOL	120	
QY	121	LVFGITANSDPHLLLOGOSLTLTLESPPSSPSVQCRSPRGNKIQQGKTLSSVQLDQSG	180	
Db	121	LVFGITANSDPHLLLOGOSLTLTLESPPSSPSVQCRSPRGNKIQQGKTLSSVQLDQSG	180	
QY	181	TWTTCTVLONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSFPLATFVEKLTGSGELMW	240	
Db	181	TWTTCTVLONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSFPLATFVEKLTGSGELMW	240	
QY	241	QAEPASSSKSWITFPLKKEVSVKATODPKLQNGKULPLHTLTPQALPQIAGSNTLTA	300	
Db	241	QAEPASSSKSWITFPLKKEVSVKATODPKLQNGKULPLHTLTPQALPQIAGSNTLTA	300	

ID	Accession	Description	Length
Db	241	Q6ERAS55SMWTFPLPKNKKEVSVKRVYQDDPKLQMGKKLPHLTLPLQALPOLYAGSGNLTLA	300
Qy	301	LEAKTGKGLHOEVNLYVMEATOLQKNLTCEVWGPTSPKLMSTLKLLENKAVSRGRXPVWY	360
Db	301	LEAKTGKGLHOEVNLYVMEATOLQKNLTCEVWGPTSPKLMSTLKLLENKAVSRGRAXVWY	360
Qy	361	LNPEAGMOCILSDSGOVLLESNTKVLPTWSTPV	394
Db	361	LNPEAGMOCILSDSGOVLLESNTKVLPTWSTPV	394
RESULT 56			
ID	Accession	Description	Length
XX	AA091369	AA091369 standard; protein; 458 AA.	
XX	AA091369;		
XX	24-OCT-2003	(revised)	
DT	25-MAR-2003	(revised)	
DT	08-JAN-1990	(first entry)	
XX	DE	T4 protein.	
XX	KM	T4 protein; human immunodeficiency virus; AIDS; clone pT4B.	
XX	OS	Homo sapiens; (human).	
XX	Key	Location/Qualifiers	
FT	Domain	1..23	
FT	Domain	24..117	
FT	Region	41	
FT	Region	109	
FT	Domain	118..134	
FT	Domain	135..189	
FT	Region	155	
FT	Region	184	
FT	Domain	190..204	
FT	Domain	205..286	
FT	Domain	287..309	
FT	Binding-site	296..298	
FT	Domain	310..376	
FT	Binding-site	325..327	
FT	Region	328	
FT	Region	370	
FT	Domain	377..397	
FT	Domain	398..420	
FT	Domain	421..458	
XX	EP330227-A.		
XX	30-AUG-1989.		
XX	24-FEB-1989;	89EP-00103297.	
XX	24-FEB-1989;	88US-00160348.	
XX	(UYCO-) COLUMBIA UNIV.		
XX	(SMYK-) SMITHKLINE BECKMAN CORP.		
XX	(UYCO) UNIV COLUMBIA NEW YORK.		
XX	(UYCO) UNIV COLUMBIA NEW YORK.		
XX	Maddon PJ, Axel R, Sweet RW, Arthos J;		
XX	WPI, 1989-250337/35.		
XX	Soluble T4 polypeptide derivs. - inhibitors of extracellular and cell to		
XX	cell spread of HIV used in prevention and treatment of AIDS.		
XX	Claim 1; Fig 6; 73pp; English.		
XX	T4 protein (AAN90619) inhibits extracellular and cell-to-cell spread of		
XX	HIV. The therapeutic agent consists of amino acids +3-+185 fused to +351-		
XX	+369; +3-+106 fused to +351-+369; or +3-+185. Also used to identify		

CC inhibitors of T4+ interactions, as target carrier proteins, and to  
 CC generate monoclonal antibodies. Above features are: Domain 1 (D); starting  
 CC at the N-terminal) = leader; D2 = variable-like-1; D3 = joining-like-1;  
 CC D4 = V2; D5 = J2; D6 = J3; D7 = V4; D8 = J4; D9 = transmembrane;  
 CC D10 = cytoplasmic; Regions are extracellular cysteines; and the 2 sites  
 CC are potential N-linked glycosylation sites. (Updated on 25-MAR-2003 to  
 CC correct PA field.) (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 458 AA;

Query Match 59.2%; Score 2021; DB 1; Length 458;  
 Best Local Similarity 99.7%; Pred. No. 3.2e-101;  
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVFPRHLLVQLALPAAATQGNKVVGKKGDVVELTCTASQKKSIOFHKNSNOIK 60  
 DB 1 MNRGVFPRHLLVQLALPAAATQGNKVVGKKGDVVELTCTASQKKSIOFHKNSNOIK 60  
 QY 61 ILGNQGSFLLTKGPSKLNDRADSRSLMDQGNFPLIIKNLIKIEDSDTYICEVEDQKEEYOL 120  
 DB 61 ILGNQGSFLLTKGPSKLNDRADSRSLMDQGNFPLIIKNLIKIEDSDTYICEVEDQKEEYOL 120  
 QY 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLSVSQLEIQDSG 180  
 DB 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLSVSQLEIQDSG 180  
 QY 181 TWICTVLQONOKVEFKIDIVLAFOKASSIVYKKEGEVFEFSPPLAFTVEKLTGSGELMW 240  
 DB 181 TWICTVLQONOKVEFKIDIVLAFOKASSIVYKKEGEVFEFSPPLAFTVEKLTGSGELMW 240  
 QY 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300  
 DB 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300  
 QY 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVWY 360  
 DB 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVWY 360  
 QY 361 LNPBAGMMQCLLSDSGQVLLSNIKVLPTWSTPV 394  
 DB 361 LNPBAGMMQCLLSDSGQVLLSNIKVLPTWSTPV 394

RESULT 57

AAI39826  
 ID AAY39826 standard; protein; 458 AA.

XX AAY39826;  
 AC  
 XX  
 DT 03-DEC-1999 (first entry)  
 XX  
 DE Soluble human T4 protein.  
 XX  
 KM Soluble T4 protein; sT4; human; HIV; binding inhibitor; T4+ cell; AIDS;  
 KM vaccine; immunisation; therapy.  
 OS Homo sapiens.  
 XX  
 OS  
 XX  
 PN US5958678-A.  
 XX  
 PD 28-SEP-1999.  
 XX  
 PF 12-DEC-1994; 94US-00354452.  
 XX  
 PR 21-AUG-1986; 86US-00898587.  
 PR 11-JUN-1991; 91US-00713564.  
 PR 06-JUL-1992; 92US-00909021.  
 XX  
 PA (UYCO ) UNIV COLUMBIA NEW YORK.  
 XX  
 PI McDougal JS, Weiss R, Axel R, Littman DR, Maddon PJ, Chess L;  
 XX WPL; 1999-561025/47.  
 DR

DR N-PSDB; AA220695.

XX Human T4 protein inhibits HIV binding to T4 cells, useful for treating  
 PT AIDS.  
 XX

XX Example 3; Fig 6; 58pp; English.

XX This sequence represents the soluble human T4 protein of the invention.  
 CC The soluble human T4 protein blocks the binding of HIV to T4+ cells and  
 CC is therefore useful for the treatment of AIDS. Monoclonal antibodies  
 CC against the T4 protein may be used as vaccines for immunising subjects  
 CC against AIDS

XX Sequence 458 AA;

Query Match 59.2%; Score 2021; DB 2; Length 458;  
 Best Local Similarity 99.7%; Pred. No. 3.2e-101;  
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MNRGVFPRHLLVQLALPAAATQGNKVVGKKGDVVELTCTASQKKSIOFHKNSNOIK 60  
 DB 1 MNRGVFPRHLLVQLALPAAATQGNKVVGKKGDVVELTCTASQKKSIOFHKNSNOIK 60  
 QY 61 ILGNQGSFLLTKGPSKLNDRADSRSLMDQGNFPLIIKNLIKIEDSDTYICEVEDQKEEYOL 120  
 DB 61 ILGNQGSFLLTKGPSKLNDRADSRSLMDQGNFPLIIKNLIKIEDSDTYICEVEDQKEEYOL 120  
 QY 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLSVSQLEIQDSG 180  
 DB 121 LVFGLTANSDTHLLOQGSITLTLESPPGSSPSVQCSPRGKNIQGGKTLSVSQLEIQDSG 180  
 QY 181 TWICTVLQONOKVEFKIDIVLAFOKASSIVYKKEGEVFEFSPPLAFTVEKLTGSGELMW 240  
 DB 181 TWICTVLQONOKVEFKIDIVLAFOKASSIVYKKEGEVFEFSPPLAFTVEKLTGSGELMW 240  
 QY 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300  
 DB 241 QABRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300  
 QY 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVWY 360  
 DB 301 LEAKTGKLEHVEVNLVVMRATOLQKNLTCEVWGPTSPKMLSLKENKEAKVSKREKPVWY 360  
 QY 361 LNPBAGMMQCLLSDSGQVLLSNIKVLPTWSTPV 394  
 DB 361 LNPBAGMMQCLLSDSGQVLLSNIKVLPTWSTPV 394

RESULT 58

AAAR04032  
 ID AAR04032 standard; protein; 2037 AA.

XX AAR04032;  
 AC  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 31-OCT-2002 (revised)  
 DT 29-MAY-1990 (first entry)  
 XX  
 DE Full length T4 encoded by plasmid pBG381.  
 XX  
 KM Soluble T4; pBG381; anti-retroviral agent; AIDS; ARC; HIV; AZT.  
 XX  
 OS Synthetic.  
 XX  
 PN WO8911860-A.  
 XX  
 PD 14-DEC-1989.  
 XX  
 PF 08-JUN-1989; 89WO-US002453.  
 XX  
 PR 10-JUN-1988; 88US-00204645.  
 PR 20-APR-1989; 89US-00341080.  
 XX



PA (BIOJ ) BIOGEN NV INC.  
 PA (GEO ) GEN HOSPITAL CORP.  
 PA (BIOJ ) BIOGEN INC.  
 PA (BIOJ ) BIOGEN INC.  
 PI Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;  
 DR WPI; 1990-007302/01.  
 DR N-PSDB; AAQ03006.  
 CC Combinations of soluble T4 protein and anti-retroviral agent - having  
 PT synergistic activity in treatment and prevention of AIDS, arc and HIV  
 PT infection.  
 PS Disclosure; Fig 2; 100pp; English.  
 CC X = stop codon. The sequence was deduced from the cDNA insert of pBG383.  
 CC Soluble T4 constructs may be produced by truncating this sequence to give  
 CC fragments from position 400 to 799, removing the transmembrane and  
 CC intracytoplasmic domains whilst retaining the extracellular region  
 CC responsible for HIV binding. The sol. T4 is combined with an anti-viral  
 CC agent such as AZT. See also AAQ03005. (updated on 31-OCT-2002 to add  
 CC missing OS field.) (Updated on 25-MAR-2003 to correct PA field.)  
 CC  
 SQ Sequence 2037 AA;  
 Query Match 59.2%; Score 2021; DB 2; Length 2037;  
 Best Local Similarity 99.7%; Pred. No. 1.5e-100;  
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MNRGVPRHLLLVQLALPAAATQGNKVLGKGGDTVELCTASQKKSIOFHMKNNOIK 60  
 DB 403 MNRGVPRHLLLVQLALPAAATQGNKVLGKGGDTVELCTASQKKSIOFHMKNNOIK 462  
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPILIKNLKIEDSDTYICEVEDQKEVQL 120  
 DB 463 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPILIKNLKIEDSDTYICEVEDQKEVQL 522  
 QY 121 LVFGLTANSDTHLLOQGSLLTLTLPSSPPSSPVQCRSPRGKNIQGGKTLVSQLELDG 180  
 DB 523 LVFGLTANSDTHLLOQGSLLTLTLPSSPPSSPVQCRSPRGKNIQGGKTLVSQLELDG 582  
 QY 181 TWCTVQLQNKVEFKIDIVLAFQKASSIVYKKEGEQVFSPLAFTVEKLTGSGELMW 240  
 DB 583 TWCTVQLQNKVEFKIDIVLAFQKASSIVYKKEGEQVFSPLAFTVEKLTGSGELMW 642  
 QY 241 QAEBASSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPOALPOYAGSGNLTIA 300  
 DB 643 QAEBASSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPOALPOYAGSGNLTIA 702  
 QY 301 LEAKTGKHOEVNLVVMRATOLQNLTCVWGPTSPKMLSLKLENKEAKVSKREKVMV 360  
 DB 703 LEAKTGKHOEVNLVVMRATOLQNLTCVWGPTSPKMLSLKLENKEAKVSKREKVMV 762  
 QY 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPWTSTPV 394  
 DB 763 LNPEAGMWQCLLSDSGQVLLSNIKVLPWTSTPV 796  
 RESULT 59  
 AAR07641 ID AAR07641 standard; protein; 2050 AA.  
 XX  
 AC AAR07641;  
 XX  
 DT 31-OCT-2002 (revised)  
 DT 20-DEC-1990 (first entry)  
 XX  
 DE Deduced sequence of pBG381 comprising truncated T4 glycoprotein.  
 XX plasmid pBG381; soluble T4 protein; AIDS; ARC; HIV.  
 XX  
 OS Synthetic.

XX Key Location/Qualifiers  
 FT 403..803  
 FT Protein /label= truncated soluble T4 glycoprotein  
 XX  
 XX W09008198-A.  
 XX  
 XX 26-JUL-1990.  
 XX  
 XX 18-JAN-1989; 89US-00300096.  
 XX  
 XX 18-JAN-1989; 89US-00300096.  
 XX  
 XX 18-JAN-1989; 89US-00300096.  
 XX  
 XX (HARD ) HARVARD COLLEGE.  
 XX  
 PI Letvin NA;  
 XX  
 DR WPI; 1990-254040/33.  
 DR N-PSDB; AAQ05608.  
 XX  
 PT Treating or preventing AIDS, ARC or HIV infection - by administering an  
 PT immunologically effective amt. of soluble T4 protein.  
 XX  
 PS Disclosure; Fig 2; 121pp; English.  
 CC  
 CC Entire sequence translation of plasmid pBG381 used to transform Chinese  
 CC Hamster Ovary cells for the production of soluble truncated T4.  
 CC Transmembrane and cytoplasmic domain-encoding regions are deleted from  
 CC the T4 CDS to encode a truncated protein. The soluble forms may be  
 CC modified to increase their immunogenicity by addition of an adjuvant such  
 CC as incomplete Freund's adjuvant. The T4 interferes with HIV/T4  
 CC interaction and elicits anti-soluble T4 antibody prodn. See also  
 CC AAQ05607. (updated on 31-OCT-2002 to add missing OS field.)  
 CC  
 SQ Sequence 2050 AA;  
 Query Match 59.2%; Score 2021; DB 2; Length 2050;  
 Best Local Similarity 99.7%; Pred. No. 1.5e-100;  
 Matches 393; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MNRGVPRHLLLVQLALPAAATQGNKVLGKGGDTVELCTASQKKSIOFHMKNNOIK 60  
 DB 403 MNRGVPRHLLLVQLALPAAATQGNKVLGKGGDTVELCTASQKKSIOFHMKNNOIK 462  
 QY 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPILIKNLKIEDSDTYICEVEDQKEVQL 120  
 DB 463 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPILIKNLKIEDSDTYICEVEDQKEVQL 522  
 QY 121 LVFGLTANSDTHLLOQGSLLTLTLPSSPPSSPVQCRSPRGKNIQGGKTLVSQLELDG 180  
 DB 523 LVFGLTANSDTHLLOQGSLLTLTLPSSPPSSPVQCRSPRGKNIQGGKTLVSQLELDG 582  
 QY 181 TWCTVQLQNKVEFKIDIVLAFQKASSIVYKKEGEQVFSPLAFTVEKLTGSGELMW 240  
 DB 583 TWCTVQLQNKVEFKIDIVLAFQKASSIVYKKEGEQVFSPLAFTVEKLTGSGELMW 642  
 QY 241 QAEBASSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPOALPOYAGSGNLTIA 300  
 DB 643 QAEBASSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPLHLTLPOALPOYAGSGNLTIA 702  
 QY 301 LEAKTGKHOEVNLVVMRATOLQNLTCVWGPTSPKMLSLKLENKEAKVSKREKVMV 360  
 DB 703 LEAKTGKHOEVNLVVMRATOLQNLTCVWGPTSPKMLSLKLENKEAKVSKREKVMV 762  
 QY 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPWTSTPV 394  
 DB 763 LNPEAGMWQCLLSDSGQVLLSNIKVLPWTSTPV 796  
 RESULT 60  
 AAP93506 ID AAP93506 standard; protein; 394 AA.  
 XX

```

AC AAP93506;
XX
XX 25-MAR-2003 (revised)
DT 02-JUN-1990 (first entry)
XX
XX Derived sequence of soluble T4 lymphocyte surface protein (sT4).
XX
XX Soluble T4 lymphocyte surface protein; sT4; AIDS therapy; AIDS diagnosis.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Protein 26..394
FT Misc-difference 26..26 /note= "When sequence was determined by amino acid
FT sequencing, this residue was 'lys.'"
FT
FT Region 27..45
FT /note= "These residues are identical to those determined
FT by amino acid sequencing"
XX
XX EPJ13377-A.
XX
XX 26-APR-1989.
XX
XX 21-OCT-1988; 88EP-00309907.
XX
XX 23-OCT-1987; 87US-00112800.
XX
XX (SMIK ) SMITHKLINE BECKMAN CORP.
XX (SMIK ) SMITHKLINE BECKMAN CORP.
XX
XX Deen KC, Folenawaas GM, Inacker RH, Sweet RW;
XX WPI; 1989-124209/17.
XX DR N-PSDB; AAN90763.
XX
XX Purifying soluble recombinant T4 lymphocyte surface protein - from cell
XX culture by adsorption on cation exchanger, elution and treatment with
XX anion exchanger.
XX
XX Disclosure; Fig 1; 13pp; English.
XX
XX The coding sequence is derived from the published sequence of sT4. sT4 is
XX useful in the prevention and treatment of AIDS by inhibiting spread of
XX the virus. It can also be used as an inhibitor of T4+ cell function, as a
XX reagent for identifying inhibitors of T4+ cell interaction and to produce
XX diagnostic monoclonal antibodies. (Updated on 25-MAR-2003 to correct PA
XX field.) (Updated on 25-MAR-2003 to correct PI field.)
XX
XX Sequence 394 AA;
XX
Query Match 59.1%; Score 2018; DB 1; Length 394;
Best Local Similarity 99.5%; Pred. No. 4e-101;
Matches 392; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
QY 1 MNRGVPRRLHLLVQLALPAATQGNKVVLGKGGDTVELCTASQKKSIOFHKNNSQIK 60
DB 1 MNRGVPRRLHLLVQLALPAATQGNKVVLGKGGDTVELCTASQKKSIOFHKNNSQIK 60
QY ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFLIKNLKTIKIDSDTYICEVEDQKEVOL 120
DB 61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNFLIKNLKTIKIDSDTYICEVEDQKEVOL 120
QY 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPKNTIOGKTLISVSOLELDQSG 180
DB 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPKNTIOGKTLISVSOLELDQSG 180
QY 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPKNTIOGKTLISVSOLELDQSG 180
DB 121 LVFGLTANSDTHLLQGQSLTLTLESPPGSSPSVQCRSPKNTIOGKTLISVSOLELDQSG 180
QY 181 TWCTCTVQONQKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLATVEKLTGSGELMW 240
DB 181 TWCTCTVQONQKVEFKIDIVLAFQKASSIVYKKEGQVEFSPPLATVEKLTGSGELMW 240
QY 241 QAERASSSKSWITFDLKNKEVSVKRVOTODPLQWGGKLPHTLTPQALPOYAGSGNLTLA 300

```

```

DB 241 QAERASSSKSWITFDLKNKEVSVKRVOTODPLQWGGKLPHTLTPQALPOYAGSGNLTLA 300
QY 301 LEAKTGKLEHVEVNLVYMRATQLOKNIICEVWGPSTSPKMLSLKLEKKAQVSRREPVMY 360
DB 301 LEAKTGKLEHVEVNLVYMRATQLOKNIICEVWGPSTSPKMLSLKLEKKAQVSRREPVMY 360
QY 361 LNPEAGMOCILSDSGQVLLSEINIKVLPWTSTPV 394
DB 361 LNPEAGMOCILSDSGQVLLSEINIKVLPWTSTPV 394

RESULT 61
AAP91922
ID AAP91922 standard; protein; 402 AA.
XX
XX AAP91922;
XX
XX 25-MAR-2003 (revised)
XX 31-OCT-2002 (revised)
DT 14-MAY-1990 (first entry)
XX
XX Sequence of a secreted form of the CD4 adhesion (CD4T) polypeptide.
XX
XX CD4 variants; CD4T; gp120; plasmid pRKCDA; HIV-1; HTLV-IIIB.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Misc-difference 25..26 /note= "signal processing site"
FT Misc-difference 366 /note= "other forms of CD4T terminate here"
FT Misc-difference 368 /note= "other forms of CD4T terminate here"
FT
FT EPJ14317-A.
XX
XX 03-MAY-1989.
XX
XX 03-OCT-1988; 88EP-00309194.
XX
XX 02-OCT-1987; 87US-00104329.
XX 28-SEP-1986; 88US-00250785.
XX
XX (GETH ) GENENTECH INC.
XX
XX Capon DJ, Gregory TJ;
XX WPI; 1989-131855/18.
XX DR N-PSDB; AAN90777.
XX
XX Compens. contg. adhesion variants - useful in therapy and diagnostics,
XX e.g. CD4 variants which are therapeutically useful for treating human
XX immuno-deficiency virus.
XX
XX Disclosure; Fig 1a-1c; 36pp; English.
XX
XX It may be capable of binding gp120. It may be fused with an
XX immunoglobulin constant domain, human transferrin, apolipoprotein,
XX albumin, ricin A chain or diphtheria toxin A. It may be used for
XX antitiral of immunomodulatory therapy particularly in treatment of HIV
XX infection. It may have variants by insertion, substitution of deletion in
XX non-functional regions. (Updated on 31-OCT-2002 to add missing OS field.)
XX (Updated on 25-MAR-2003 to correct PR field.) (Updated on 25-MAR-2003 to
XX correct PI field.)
XX
XX Sequence 402 AA;
XX
Query Match 59.1%; Score 2017; DB 1; Length 402;
Best Local Similarity 99.7%; Pred. No. 4.7e-101;
Matches 392; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 MNRGVPRRLHLLVQLALPAATQGNKVVLGKGGDTVELCTASQKKSIOFHKNNSQIK 60

```

```

Db      1  MNRGVPRHLLLVQLALLPAATQGNKVLGKKGDVLELCTASQKKSIOFHMKNQIK 60
Qy      61  ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120
Db      61  ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120
Qy      121  LVFGLTANSSTHLLQGSGLTLTLSPSSPVOCRSRPGKNIQSGKTLVSQLELDSG 180
Db      121  LVFGLTANSSTHLLQGSGLTLTLSPSSPVOCRSRPGKNIQSGKTLVSQLELDSG 180
Qy      181  TWTCVTVLONQKVEFKIDIVLAFQKASSIVYKKEGEVFPPLAFTVEKLTGSGELMW 240
Db      181  TWTCVTVLONQKVEFKIDIVLAFQKASSIVYKKEGEVFPPLAFTVEKLTGSGELMW 240
Qy      241  QAERASSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300
Db      241  QAERASSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300
Qy      301  LEAKTGKLEHENVLVVWRATOLQKNTLCEVWGPTSPKMLSLKLENKEAVSKREKPVWV 360
Db      301  LEAKTGKLEHENVLVVWRATOLQKNTLCEVWGPTSPKMLSLKLENKEAVSKREKPVWV 360
Qy      361  LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTP 393
Db      361  LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTP 393

```

## RESULT 62

AAP94757 ID AAP94757 standard; protein; 402 AA.

XX AAP94757;

DT 25-MAR-2003 (revised)  
 DT 03-OCT-2002 (revised)  
 DT 28-JAN-1991 (first entry)

DE Sequence of a secreted form of the CD4 adhesion.

XX HIV; antiviral; therapy; diagnosis.

XX Homo sapiens.

XX Key location/Qualifiers

FT Peptide 1..25 /note="signal"

FT Protein 26..402

XX MO8902922-A.

XX 06-APR-1989.

XX 03-OCT-1988; 88WO-US003414.

XX 02-OCT-1987; 87US-00104329.

XX 28-SEP-1986; 88US-00250785.

XX (GETH ) GENENTECH INC.

XX Capon DJ, Gregory TJ;

XX WPI; 1989-114397/15.

XX N-PSDB; AAN90734.

XX New nucleic acid sequences encoding adhesion, esp. CD 4, variants -

XX partic. with trans-membrane domain inactivated or fused to other peptide,

XX useful esp. for treating HIV infections.

XX Disclosure; Fig 1a-1c; 78pp; English.

XX The patent claims a nucleic acid encoding an aa sequence variant of an

XX adhesion, which is pref. a CD4 polypeptide variant modified such that its

CC transmembrane domain has been inactivated, either deleted or replaced by  
 CC a sequence of hydrophilic hydropathy profile. The aa sequence variant of  
 CC an adhesion may also be a fusion of CD4 with a 2nd polypeptide esp. one  
 CC contg. a non-CD4 epitope; a signal sequence; a cgd. able to elicit a  
 CC humoral immune response (viral polypeptide or allergen); or a human  
 CC plasma protein of long plasma half-life. CD4 fusion proteins can have  
 CC antiviral and immunomodulatory activity and are esp. useful for treating  
 CC HIV infections regardless of genetic variation within the virus. They and  
 CC antibodies raised against them can also be used diagnostically for  
 CC assaying adhesions and their ligands. (updated on 03-OCT-2002 to add  
 CC missing OS field.) (updated on 25-MAR-2003 to correct PR field.) (updated  
 CC on 25-MAR-2003 to correct PA field.)

XX Sequence 402 AA;

Query Match 59.1%; Score 2017; DB 1; Length 402;  
 Best Local Similarity 99.7%; Pred. No. 4,7e-101;  
 Matches 392; Conservative 1; Indels 0; Gaps 0;

```

Qy      1  MNRGVPRHLLLVQLALLPAATQGNKVLGKKGDVLELCTASQKKSIOFHMKNQIK 60
Db      1  MNRGVPRHLLLVQLALLPAATQGNKVLGKKGDVLELCTASQKKSIOFHMKNQIK 60
Qy      61  ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120
Db      61  ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQL 120
Qy      121  LVFGLTANSSTHLLQGSGLTLTLSPSSPVOCRSRPGKNIQSGKTLVSQLELDSG 180
Db      121  LVFGLTANSSTHLLQGSGLTLTLSPSSPVOCRSRPGKNIQSGKTLVSQLELDSG 180
Qy      181  TWTCVTVLONQKVEFKIDIVLAFQKASSIVYKKEGEVFPPLAFTVEKLTGSGELMW 240
Db      181  TWTCVTVLONQKVEFKIDIVLAFQKASSIVYKKEGEVFPPLAFTVEKLTGSGELMW 240
Qy      241  QAERASSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300
Db      241  QAERASSSKSWITFDLKNKEVSVKRVTPDPLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300
Qy      301  LEAKTGKLEHENVLVVWRATOLQKNTLCEVWGPTSPKMLSLKLENKEAVSKREKPVWV 360
Db      301  LEAKTGKLEHENVLVVWRATOLQKNTLCEVWGPTSPKMLSLKLENKEAVSKREKPVWV 360
Qy      361  LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTP 393
Db      361  LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTP 393

```

## RESULT 63

AA88328 ID AA88328 standard; protein; 394 AA.

XX AA88328;

DT 14-JUL-2000 (first entry)

DE T4 glycoprotein amino acid sequence.

XX sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;

XX AIDS; treatment; inhibit; cell to cell spread; infection; fusion.

XX Mammalia.

XX US5126433-A.

XX 30-JUN-1992.

XX 23-OCT-1987; 87US-00114244.

XX 21-AUG-1986; 86US-00898587.

XX (UYCO ) UNIV COLUMBIA NEW YORK.

PI Madden PJ, Chess L, Axel R, Weiss R, Littman DR, McDougal JS;  
 XX MPI; 2000-348913/30.  
 DR N-PSDB; AAA10906.  
 XX  
 XX Soluble T4 glycoprotein useful for prevention and treatment of acquired  
 PT immunodeficiency syndrome and for screening inhibitors of human  
 PT immunodeficiency viral binding.  
 XX  
 XX Disclosure; Col 11-16; 64pp; English.  
 XX  
 XX This sequence represents the full length amino acid sequence of  
 CC glycosylated sT4 glycoprotein. Human immunodeficiency virus (HIV) uses  
 CC sT4 as a target receptor on T cells. The invention relates to  
 CC glycosylated sT4 which functions by blocking the binding of HIV to T4  
 CC target cells, and can be used for the prophylaxis and treatment of AIDS  
 CC patients. Administration of sT4 effectively inhibits the cell to cell  
 CC spreading of HIV infection and also the fusion of HIV-infected T4 cells  
 CC and non-infected T4 cells. The administration of T4 alleviates several  
 CC symptoms associated with AIDS, and prevents the occurrence of new  
 CC pathological changes. The sT4 glycoprotein is useful for the prophylaxis  
 CC and treatment of patients with AIDS. It is also useful as a reagent to  
 CC identify natural, synthetic or recombinant molecules which act as  
 CC therapeutic agents or inhibitors of T4 cell interactions and in  
 CC diagnostic assays for detection of T4 proteins or molecules  
 CC  
 XX Sequence 394 AA;  
 SQ  
 Query Match 59.0%; Score 2015; DB 3; Length 394;  
 Best Local Similarity 99.5%; Pred. No. 5.8e-101;  
 Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MNRGVPRHLLVQLALPAATQGNKVLGKKGDVELTCTASOKKSIOFHKNSNOIK 60  
 DB 1 MNRGVPRHLLVQLALPAATQGNKVLGKKGDVELTCTASOKKSIOFHKNSNOIK 60  
 QY 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNPFLITKNLKIEDSDTYICEVEDQKEEVL 120  
 DB 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNPFLITKNLKIEDSDTYICEVEDQKEEVL 120  
 QY 121 LVFGLTANSDTHLLOQSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDTHLLOQSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWTCVTLOQOKKVEFKIDIVLAFOKASSIVYKKEEQVEFSPLAFTYEKLTGSGELMW 240  
 DB 181 TWTCVTLOQOKKVEFKIDIVLAFOKASSIVYKKEEQVEFSPLAFTYEKLTGSGELMW 240  
 QY 241 QAEKASSSKSWITFDLKNKEVSVKRVTPQPKLOMGKCLPLHLTLPOLPOYAGSGNLTLLA 300  
 DB 241 QAEKASSSKSWITFDLKNKEVSVKRVTPQPKLOMGKCLPLHLTLPOLPOYAGSGNLTLLA 300  
 QY 301 LEAKTGKLGHEVNLVVMRAATQLOKNLTCEWGPSTPKMLSLKLENKEAKVSRREKPVVW 360  
 DB 301 LEAKTGKLGHEVNLVVMRAATQLOKNLTCEWGPSTPKMLSLKLENKEAKVSRREKPVVW 360  
 QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394  
 DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394  
 RESULT 64  
 ID AAY88329 standard; protein; 458 AA.  
 XX  
 XX AAY88329;  
 XX  
 DT 14-JUL-2000 (first entry)  
 XX  
 XX T4 glycoprotein amino acid sequence.  
 DE  
 XX sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;  
 KM AIDS; treatment; inhibit; cell to cell spread; infection; fusion.

XX  
 OS Mammalia.  
 XX  
 PN US5126433-A.  
 XX  
 PD 30-JUN-1992.  
 XX  
 XX 23-OCT-1987; 87US-00114244.  
 PF  
 XX 21-AUG-1986; 86US-00898587.  
 PR  
 XX (UYCO ) UNIV COLUMBIA NEW YORK.  
 PA  
 PI Madden PJ, Chess L, Axel R, Weiss R, Littman DR, McDougal JS;  
 XX MPI; 2000-348913/30.  
 DR  
 XX Soluble T4 glycoprotein useful for prevention and treatment of acquired  
 PT immunodeficiency syndrome and for screening inhibitors of human  
 PT immunodeficiency viral binding.  
 XX  
 XX Example; Fig 6; 64pp; English.  
 XX  
 XX This sequence represents the amino acid sequence of glycosylated sT4  
 CC glycoprotein. Human immunodeficiency virus (HIV) uses sT4 as a target  
 CC receptor on T cells. The invention relates to glycosylated sT4 which  
 CC functions by blocking the binding of HIV to T4 target cells, and can be  
 CC used for the prophylaxis and treatment of AIDS patients. Administration  
 CC of sT4 effectively inhibits the cell to cell spreading of HIV infection  
 CC and also the fusion of HIV-infected T4 cells and non-infected T4 cells.  
 CC The administration of T4 alleviates several symptoms associated with  
 CC AIDS, and prevents the occurrence of new pathological changes. The sT4  
 CC glycoprotein is useful for the prophylaxis and treatment of patients with  
 CC AIDS. It is also useful as a reagent to identify natural, synthetic or  
 CC recombinant molecules which act as therapeutic agents or inhibitors of  
 CC T4 cell interactions and in diagnostic assays for detection of T4 proteins  
 CC or molecules  
 CC  
 XX Sequence 458 AA;  
 SQ  
 Query Match 59.0%; Score 2015; DB 3; Length 458;  
 Best Local Similarity 99.5%; Pred. No. 6.8e-101;  
 Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;  
 QY 1 MNRGVPRHLLVQLALPAATQGNKVLGKKGDVELTCTASOKKSIOFHKNSNOIK 60  
 DB 1 MNRGVPRHLLVQLALPAATQGNKVLGKKGDVELTCTASOKKSIOFHKNSNOIK 60  
 QY 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNPFLITKNLKIEDSDTYICEVEDQKEEVL 120  
 DB 61 ILGNQGSFLTKGPSKLNDRADRSRLMDQGNPFLITKNLKIEDSDTYICEVEDQKEEVL 120  
 QY 121 LVFGLTANSDTHLLOQSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDTHLLOQSILTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWTCVTLOQOKKVEFKIDIVLAFOKASSIVYKKEEQVEFSPLAFTYEKLTGSGELMW 240  
 DB 181 TWTCVTLOQOKKVEFKIDIVLAFOKASSIVYKKEEQVEFSPLAFTYEKLTGSGELMW 240  
 QY 241 QAEKASSSKSWITFDLKNKEVSVKRVTPQPKLOMGKCLPLHLTLPOLPOYAGSGNLTLLA 300  
 DB 241 QAEKASSSKSWITFDLKNKEVSVKRVTPQPKLOMGKCLPLHLTLPOLPOYAGSGNLTLLA 300  
 QY 301 LEAKTGKLGHEVNLVVMRAATQLOKNLTCEWGPSTPKMLSLKLENKEAKVSRREKPVVW 360  
 DB 301 LEAKTGKLGHEVNLVVMRAATQLOKNLTCEWGPSTPKMLSLKLENKEAKVSRREKPVVW 360  
 QY 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394  
 DB 361 LNPEAGMOCCLSDSGOVLLESNIKVLPTWSTPV 394

ID	AAB81502	standard; protein; 458 AA.
XX	AAB81502;	
AC	AAB81502;	
DT	18-JUN-2001	(first entry)
XX		
DE	Human CD4 protein.	
XX		
KM	Human; CD4; CD4 fusion protein; oligomerisation; receptor-ligand interaction inhibition; surface plasmon resonance; SPR; T cell receptor binding; MHC binding; carcinoma; autoimmune disease; multiple sclerosis; human immunodeficiency virus; HIV; diabetes; rheumatoid arthritis; immune disorder.	
XX		
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..25
FT	Protein	/label= Signal_peptide 26..458 /label= Human_CD4
XX		
PN	WO200122084-A2.	
XX		
PD	29-MAR-2001.	
XX		
PF	18-SEP-2000; 2000MO-GB003579.	
XX		
PR	21-SEP-1999; 99GB-00022352.	
PA	(AVID-) AVIDEX LTD.	
PI	Jakobsen BK;	
DR	WI: 2001-273470/28.	
DR	N-PSDB; AAF82582.	
PT	Sequential screening of candidate compounds library for those which inhibit binding of low affinity receptor-ligand interaction having fast binding kinetics, using interfacial optical assay.	
PS	Disclosure; Fig 13; 91pp; English.	
XX		
CC	The present sequence is human CD4. Human CD4 extracellular domains 1 and 2 were used in the construction of CD4 oligomerisation fusion proteins. The fusion proteins contain an oligomerisation domain that enables the proteins to bind to one another to form oligomers. The oligomers may be used in an invention relating to a method for screening for compounds with the ability to inhibit a low affinity receptor-ligand interaction. The method uses an interfacial optical assay, such as surface plasmon resonance (SPR). The method is useful for screening candidate compounds for the ability to inhibit interaction between MHC/peptide complex and T cell receptor, and MHC/peptide complex and CD8 or CD4 co-receptor. The compound identified by the above methods which interfere with T cell receptor binding to a particular HLA type molecule are useful as immune inhibitors for treating carcinomas, autoimmune diseases such as immune sclerosis, human immunodeficiency virus (HIV) infection, rheumatoid arthritis, Hansen's disease, insulin dependent diabetes, Good pasture's syndrome, uveitis, psoriasis and graft rejection	
XX		
SQ	Sequence 458 AA;	
Query Match	59.0%; Score 2015; DB 4; Length 458;	
Best Local Similarity	99.5%; Pred. No. 6,8e-101; Indels 0; Gaps 0;	
Matches 392; Conservative 0; Mismatches 2;		
D5	1 MNRGVPPRHLLLVLTQALLPAAQONKVVLGKKGDVEVLGTASQKSIOFHMKNSNOIK 60	
D6	1 MNRGVPPRHLLLVLTQALLPAAQOKVVVGKGDVEVLGTASQKSIOFHMKNSNOIK 60	
D7	61 ILGNQGSFLTGPKSLANDRADSRNLSMDQGNPLIINKLKIEDSDTYICEVEDQKEVOL 120	

Db	61	ILGNQGSFLITGSPSKLNDRADRSKSLMDQGNFPLITIKNLKIEDSDTYICEVEDQKEEYOL	120
Qy	121	LVFGITANSDFHLIQGSLTTLTLESPPGSSPSVOCRSBRGNKIQGKTLVSQLELODSG	180
Db	121	LVFGITANSDFHLIQGSLTTLTLESPPGSSPSVOCRSBRGNKIQGKTLVSQLELODSG	180
Qy	181	TWCTCTVLONQKKVEPKDIDVILVAFQAKSIYKKKEGEQVEFSPLAFTVEKLTGSGELMW	240
Db	181	TWCTCTVLONQKKVEPKDIDVILVAFQAKSIYKKKEGEQVEFSPLAFTVEKLTGSGELMW	240
Qy	241	QAEBSASSKSKITFDLKNKKEVSVRKVPDPLQMGKTLPLHLTLPOLPYAGSGNLTLA	300
Db	241	QAEBSASSKSKITFDLKNKKEVSVRKVPDPLQMGKTLPLHLTLPOLPYAGSGNLTLA	300
Qy	301	LEAKTGKLGHOEVNLVVMARATOLQNLTCENVGPPSPKLMSTLKLLENKAYSKKEKPYWV	360
Db	301	LEAKTGKLGHOEVNLVVMARATOLQNLTCENVGPPSPKLMSTLKLLENKAYSKKEKPYWV	360
Qy	361	LNPEAGMWQCLLSDSGOVLLESNIKVLPTWSTPV	394
Db	361	LNPEAGMWQCLLSDSGOVLLESNIKVLPTWSTPV	394
RESULT 66			
ADD25609	ID	ADD25609 standard; protein; 458 AA.	
XX	AC	ADD25609;	
XX	DT	15-JAN-2004 (first entry)	
DE	XX	Binding domain-immunoglobulin fusion protein-associated protein #82.	
KM	XX	Binding domain; immunoglobulin; fusion protein; cytostatic;	
KM	XX	antiarthritic; immunosuppressive; antidiabetic; antihypoid;	
KM	XX	neuroprotective; hinge region; immunoglobulin heavy chain;	
KM	XX	CH2 constant region; CH3 constant region; IgG1;	
KM	XX	antibody dependent cell-mediated cytotoxicity; ADCC; complement fixation;	
KM	XX	malignant condition; B-cell disorder; melanoma; carcinoma; sarcoma;	
KM	XX	rheumatoid arthritis; myasthenia gravis; Grave's disease;	
XX	XX	type I diabetes mellitus; multiple sclerosis; autoimmune disease.	
OS	XX	unidentified.	
XX	XX	US2003118592-A1.	
PN	XX	26-JUN-2003.	
PD	XX	25-JUL-2002; 2002US-00207655.	
PF	XX	17-JAN-2001; 2001US-0367358P.	
PR	XX	17-JAN-2002; 2002US-00053530.	
PR	XX	03-JUN-2002; 2002US-0385691P.	
XX	PA	(GENE-) GENECLRAFT INC.	
XX	PI	Ledbetter JA, Hayden-Ledbetter MS, Thompson PA;	
XX	PI	WPI; 2003-801317/75.	
DR	XX	New binding domain-immunoglobulin fusion protein, useful for treating a	
PT	XX	subject having or suspected of having a malignant condition or a B-cell	
XX	XX	disorder, e.g. melanoma, Grave's disease or autoimmune disease.	
PS	XX	Disclosure; SEQ ID NO 170; 157pp; English.	
CC	XX	The invention relates to a binding domain-immunoglobulin fusion protein	
CC	XX	comprising a binding domain polypeptide that is fused to an	
CC	XX	immunoglobulin hinge region polypeptide, an immunoglobulin heavy chain	
CC	XX	CH2 constant region polypeptide that is fused to the hinge region	
CC	XX	polypeptide, and an immunoglobulin heavy chain CH3 constant region	
CC	XX	polypeptide that is fused to the CH2 constant region polypeptide. The	

CC hinge region polypeptide comprises: a wild-type human IgG1 immunoglobulin  
CC hinge region polypeptide; a mutated human IgG1 immunoglobulin hinge  
CC region polypeptide, derived from (a) having 3 or more cysteine residues;  
CC where the mutated human IgG1 immunoglobulin hinge region polypeptide  
CC contains 2 cysteine residues, where the first cysteine is not mutated; a  
CC mutated human IgG1 immunoglobulin hinge region polypeptide, derived from  
CC (a) having 3 or more cysteine residues, where the mutated human IgG1  
CC immunoglobulin hinge region polypeptide contains no more than one  
CC cysteine residue; and a mutated human IgG1 immunoglobulin hinge region  
CC polypeptide, derived from (a) having 3 or more cysteine residues; where  
CC the mutated human IgG1 immunoglobulin hinge region polypeptide contains  
CC no cysteine residues. The binding domain-immunoglobulin fusion protein is  
CC capable of at least one immunological activity comprising antibody  
CC dependent cell-mediated cytotoxicity (ADCC) and complement fixation. The  
CC binding domain polypeptide is capable of specifically binding to an  
CC antigen. Also included are an isolated polynucleotide encoding the  
CC binding domain-immunoglobulin fusion protein, a recombinant expression  
CC construct comprising the polynucleotide (operably linked to a promoter),  
CC a host cell transformed or transfected with a recombinant expression  
CC construct, producing the binding domain-immunoglobulin fusion protein, a  
CC pharmaceutical composition comprising the binding domain-immunoglobulin  
CC fusion protein or polynucleotide and a carrier, and treating a subject  
CC having or suspected of having a malignant condition or a B-cell disorder.  
CC The binding domain-immunoglobulin fusion protein is useful for treating a  
CC subject having or suspected of having a malignant condition or a B-cell  
CC disorder, e.g. melanoma, carcinoma or sarcoma, rheumatoid arthritis,  
CC myasthenia gravis, Grave's disease, type 1 diabetes mellitus, multiple  
CC sclerosis or autoimmune disease. The present sequence is a binding domain  
CC -immunoglobulin fusion protein-associated protein sequence. Note: The  
CC sequence data for this patent formed part of the printed specification  
CC and is also available in electronic format directly from USPTO at  
CC seqdata.uspto.gov/sequence.html?docid=20030118592. The authors have not  
CC identified the sequences in the printed specification by their SEQ ID  
CC number therefore none of the sequences can be explicitly identified.

SQ Sequence 458 AA;

Query Match 59.0%; Score 2015; DB 7; Length 458;

Best Local Similarity 99.5%; Pred. No. 6.8e-101;

Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVFRRHLLVLTQALLPAATQGNKVVYLGKKGDVVELTCTASQKKSIOFHKNSNQIK 60  
DB 1 MNRGVFRRHLLVLTQALLPAATQGNKVVYLGKKGDVVELTCTASQKKSIOFHKNSNQIK 60  
QY 61 ILGNQGSFLTKGSKLNDRAADSRSLMDQGNFLLIINKLIEDSDTYICEVEDQKEVOL 120  
DB 61 ILGNQGSFLTKGSKLNDRAADSRSLMDQGNFLLIINKLIEDSDTYICEVEDQKEVOL 120  
QY 121 LVFGLTANSPTHLLOQGSLLTLTPSPGSSPSVQCRRPKKNIOGGKTLSSVSLQLQDSG 180  
DB 121 LVFGLTANSPTHLLOQGSLLTLTPSPGSSPSVQCRRPKKNIOGGKTLSSVSLQLQDSG 180  
QY 181 TWICTTVLQNGKXVEFKIDIVLAFOKASIVYKKEGQVFFSPLAFTVEKLTGSGELMW 240  
DB 181 TWICTTVLQNGKXVEFKIDIVLAFOKASIVYKKEGQVFFSPLAFTVEKLTGSGELMW 240  
QY 241 QABRASSSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPHLTLPLQALPOYAGSGNITLA 300  
DB 241 QABRASSSKSWITFDLKNKEVSVKRVTDQPKLQMGKKLPHLTLPLQALPOYAGSGNITLA 300  
QY 301 LEAKTGLHOEVNLVYWRARATQLOKNLTCEVWGPSPKLMSTKLKNEAASVREKRVWV 360  
DB 301 LEAKTGLHOEVNLVYWRARATQLOKNLTCEVWGPSPKLMSTKLKNEAASVREKRVWV 360  
QY 361 LNPEAGWQCLISDSGOVLLESNIKVLPTWSTPV 394  
DB 361 LNPEAGWQCLISDSGOVLLESNIKVLPTWSTPV 394

RESULT 67  
ADE57489  
ID ADE57489 standard; protein: 458 AA.

XX ADE57489;  
AC 29-JAN-2004 (first entry)  
DT Human Protein P01730, SEQ ID NO 3351.  
DE Human Protein P01730, SEQ ID NO 3351.  
KW Human; pain; neuronal tissue; gene therapy;  
KW spinal segmental nerve injury; chronic constriction injury; CCI;  
KW spared nerve injury; SNI; Chung.  
XX Homo sapiens.  
XX MO2003016475-A2.  
XX 27-FEB-2003.  
XX 14-AUG-2002; 2002WO-US025765.  
XX 14-AUG-2001; 2001US-0312147P.  
XX 01-NOV-2001; 2001US-0346382P.  
XX 26-NOV-2001; 2001US-033347P.  
XX (GENO ) GEN HOSPITAL CORP.  
XX (FARB ) BAYER AG.  
XX Woolf C, D'ureo D, Befort K, Costigan M;  
XX WPI; 2003-268312/26.  
XX GENBANK; P01730.  
XX New composition comprising two or more isolated polypeptides, useful for  
XX preparing a medicament for treating pain in an animal.  
XX Claim 1; Page: 1017pp; English.  
XX The invention discloses a composition comprising two or more isolated rat  
XX or human polynucleotides or a polynucleotide which represents a fragment,  
XX derivative or allelic variation of the nucleic acid sequence. Also  
XX claimed are a vector comprising the novel polynucleotide, a host cell  
XX comprising the vector, a method for identifying a nucleotide sequence  
XX which is differentially regulated in an animal subjected to pain and a  
XX kit to perform the method, an array, a method for identifying an agent  
XX that increases or decreases the expression of the polynucleotide sequence  
XX that is differentially expressed in neuronal tissue of a first animal  
XX subjected to pain, a method for identifying a compound which regulates  
XX the expression of a polynucleotide sequence which is differentially  
XX expressed in an animal subjected to pain, a method for identifying a  
XX compound that regulates the activity of one or more of the  
XX polynucleotides, a method for producing a pharmaceutical composition, a  
XX method for identifying a compound or small molecule that regulates the  
XX activity in an animal of one or more of the polypeptides given in the  
XX specification, a method for identifying a compound useful in treating  
XX pain and a pharmaceutical composition comprising the one or more  
XX polypeptides or their antibodies. The polynucleotide or the compound that  
XX modulates its activity is useful for preparing a medicament for treating  
XX pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
XX injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
XX therapy). The sequence presented is a human protein (shown in Table 2 of  
XX the specification) which is differentially expressed during pain. Note:  
XX The sequence data for this patent did not form part of the printed  
XX specification, but was obtained in electronic form directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences.

SQ Sequence 458 AA;

Query Match 59.0%; Score 2015; DB 7; Length 458;

Best Local Similarity 99.5%; Pred. No. 6.8e-101;

Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVFRRHLLVLTQALLPAATQGNKVVYLGKKGDVVELTCTASQKKSIOFHKNSNQIK 60  
DB 1 MNRGVFRRHLLVLTQALLPAATQGNKVVYLGKKGDVVELTCTASQKKSIOFHKNSNQIK 60

QY 61 ILGNQGSFLTKGPKLNDNRADSRSLMDQGNFPLIKKLIKEDSDTYICEVEDQKEEVOL 120  
 DB 61 ILGNQGSFLTKGPKLNDNRADSRSLMDQGNFPLIKKLIKEDSDTYICEVEDQKEEVOL 120  
 QY 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVOCSPRGKNIQGGKTLVSQLELQDSG 180  
 DB 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVOCSPRGKNIQGGKTLVSQLELQDSG 180  
 QY 181 TWCTTVLQONKQKVEFKIDIVVLAFOKASSIYKKEGEQVEFSPPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTTVLQONKQKVEFKIDIVVLAFOKASSIYKKEGEQVEFSPPLAFTVEKLTGSGELMW 240  
 QY 241 QAEBSASSKSWITFDLKNKEVSVRVYTOPDKLQMGKPLPHLTLPOALPOYAGSGNLTIA 300  
 DB 241 QAEBSASSKSWITFDLKNKEVSVRVYTOPDKLQMGKPLPHLTLPOALPOYAGSGNLTIA 300  
 QY 301 LEAKTGKLGHEVNLVVRATOLQKNLTCEVWGPTSPKLMSTLKENKEAKVSKREKAPVW 360  
 DB 301 LEAKTGKLGHEVNLVVRATOLQKNLTCEVWGPTSPKLMSTLKENKEAKVSKREKAPVW 360  
 QY 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394  
 DB 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394

## RESULT 68

ADA44807  
 ID ADA44807 standard; protein; 473 AA.

AC ADA44807;

DT 04-DEC-2003 (first entry)

DE CD4/TCR CD3epsilon chain chimeric protein CD4epsilon10n5, SEQ ID NO:2.

KM HIV-1 infection; human immunodeficiency virus-1; CD4+ cell; chimeric CD4;

KM endoplasmic reticulum; ER retention; envelope protein gp160;

KM T cell receptor CD3epsilon chain; C-terminal domain; CD4epsilon10n5;

KM gene therapy; human; receptor.

OS Chimeric.

OS Homo sapiens.

FT Key Location/Qualifiers

FT Protein 1..458

FT Region /label= CD4

FT 459..473

FT N-terminus "Part of the C-terminal domain of the T cell

FT receptor CD3epsilon chain"

FT 459..473

FT 459..473

FT 459..473

FT 459..473

FT 459..473

FT 459..473

FT 459..473

CC The invention relates to a composition for the treatment or prevention of  
 CC human immunodeficiency virus-1 (HIV-1) infection. The composition  
 CC comprises CD4+ cells that have been transduced with a vector that encodes  
 CC a chimeric CD4 molecule which is capable of being retained in the  
 CC endoplasmic reticulum (ER). The invention also encompasses the use of a  
 CC soluble protein factor produced by CD4+ cells that have been transduced  
 CC with a vector encoding a chimeric CD4 protein; and the use of an  
 CC expression system encoding a chimeric CD4 protein. The ER-localised  
 CC chimeric CD4 molecule binds to the HIV-1 envelope protein gp160,  
 CC resulting in HIV-1 retention in the ER and thereby preventing viral  
 CC replication. In a specific embodiment, the chimeric CD4 molecule  
 CC comprises CD4 fused to 15 amino acids of the C-terminal domain of the T  
 CC cell receptor CD3epsilon chain; this chimeric CD4 molecule is designated  
 CC CD4epsilon10n5 (ADA44807). A known chimeric CD4 of similar structure but  
 CC containing only 10 amino acids from CD3epsilon can also be used.  
 CC Compositions of the invention have an in trans effect on the replication  
 CC of HIV-1, and may be used to treat and prevent HIV-1 infection. The  
 CC present sequence represents the chimeric CD4 molecule CD4epsilon10n5, which  
 CC is specifically claimed for use in compositions of the invention.

XX Sequence 473 AA;

Query Match 59.0%; Score 2015; DB 7; Length 473;

Best Local Similarity 99.5%; Pred. No. 7.1e-101;

Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTQLALPAATQGNKVYLGKGDVYELTCTASQKKSIOFHMKNQIK 60  
 DB 1 MNRGVPRHLLVLTQLALPAATQGNKVYLGKGDVYELTCTASQKKSIOFHMKNQIK 60  
 QY 61 ILGNQGSFLTKGPKLNDNRADSRSLMDQGNFPLIKKLIKEDSDTYICEVEDQKEEVOL 120  
 DB 61 ILGNQGSFLTKGPKLNDNRADSRSLMDQGNFPLIKKLIKEDSDTYICEVEDQKEEVOL 120  
 QY 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVOCSPRGKNIQGGKTLVSQLELQDSG 180  
 DB 121 LVFGLTANSDTHLLQGGSLTTLTSPGSSPSVOCSPRGKNIQGGKTLVSQLELQDSG 180  
 QY 181 TWCTTVLQONKQKVEFKIDIVVLAFOKASSIYKKEGEQVEFSPPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTTVLQONKQKVEFKIDIVVLAFOKASSIYKKEGEQVEFSPPLAFTVEKLTGSGELMW 240  
 QY 241 QAEBSASSKSWITFDLKNKEVSVRVYTOPDKLQMGKPLPHLTLPOALPOYAGSGNLTIA 300  
 DB 241 QAEBSASSKSWITFDLKNKEVSVRVYTOPDKLQMGKPLPHLTLPOALPOYAGSGNLTIA 300  
 QY 301 LEAKTGKLGHEVNLVVRATOLQKNLTCEVWGPTSPKLMSTLKENKEAKVSKREKAPVW 360  
 DB 301 LEAKTGKLGHEVNLVVRATOLQKNLTCEVWGPTSPKLMSTLKENKEAKVSKREKAPVW 360  
 QY 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394  
 DB 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPV 394

## RESULT 69

AAR20152  
 ID AAR20152 standard; protein; 519 AA.

AC AAR20152;

DT 25-MAR-2003 (revised)

DT 31-MAR-1992 (first entry)

DE Human CD4 sequence encoded by PATY. 6.

KM Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;

KM acquired immune deficiency syndrome; AIDS related complex;

OS Homo sapiens.

FT Key Location/Qualifiers

```

FT Peptide 1..25
FT /label= signal_sequence
XX WO9118618-A.
XX 12-DEC-1991.
XX
XX 25-MAY-1990; 90US-00529186.
XX
XX 25-MAY-1990; 90US-00529186.
XX
XX PR 25-MAY-1990; 90US-00529186.
XX
XX PA (BIOU ) BIOGEN INC.
XX
XX PI Fisher RA, Hession C, Burkly LC;
XX
XX DR WPI, 1992-007200/01.
XX
XX DR N-PSDB; AAQ20327.
XX
XX PT New immuno-therapeutic human CD4 variants and deriva. - eliciti AB
XX PT production to HIV gp120, useful in treating, preventing and diagnosing
XX PT AIDS, ARC and HIV infections.
XX
XX PS Disclosure; Fig 28; 179pp; English.
XX
XX CC The sequence was deduced from the DNA sequence of subclone PATY.6, contg.
XX CC DNA coding for the full-length human CD4. The clone was constructed from
XX CC plasmids pBG178A and pBG378 (both in US8802940). The DNA can be used to
XX CC express recombinant CD4 and analogues for use in diagnosis and treatment
XX CC of diseases caused by infective agents whose primary targets are T4+
XX CC lymphocytes. See also AAR20148-R20155 and AAR21078. (Updated on 25-MAR-
XX CC 2003 to correct PA field.)
XX
XX SQ Sequence 519 AA;

Query Match 59.0%; Score 2015; DB 2; Length 519;
Best Local Similarity 99.5%; Pred. No. 7.8e-101;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTQLALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 60
DB 62 MNRGVPRHLLVLTQLALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 121
QY 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLKIENSDTYICEVEDQKEVOL 120
DB 122 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLKIENSDTYICEVEDQKEVOL 181
QY 121 LVFGLTANSDTHLLQGSLLTLESPPSSPSVQCRSPRGKNTQGGKTLVSQLELDQSG 180
DB 182 LVFGLTANSDTHLLQGSLLTLESPPSSPSVQCRSPRGKNTQGGKTLVSQLELDQSG 241
QY 181 TWICTTVLQONKVFVKIDIVVLAFOKASSIVYKKEGVQVFPPLATTVTEKLTGSGELMW 240
DB 242 TWICTTVLQONKVFVKIDIVVLAFOKASSIVYKKEGVQVFPPLATTVTEKLTGSGELMW 301
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTPDPLQMGKTLPHLTLTPOALPOYASSGNLTLLA 300
DB 302 QAERASSSKSWITFDLKNKEVSVKRVTPDPLQMGKTLPHLTLTPOALPOYASSGNLTLLA 361
QY 301 LEAKTGLHQBVLVVMRATQLOKNTLCEVWGPTSPKLMLSLKLENKAVSKREKRVW 360
DB 362 LEAKTGLHQBVLVVMRATQLOKNTLCEVWGPTSPKLMLSLKLENKAVSKREKRVW 421
QY 361 LNPAGMWOCILSDSGVLLSNIKVLPWTSTPV 364
DB 422 LNPAGMWOCILSDSGVLLSNIKVLPWTSTPV 455

```

RESULT 70

```

AAR13491
ID AAR13491 standard; protein; 458 AA.
XX
XX AC AAR13491;
XX

```

```

DT 25-MAR-2003 (revised)
DT 30-OCT-1991 (first entry)
XX
XX DE Human CD4 encoded by pJOD.sCD4.Y187.Snab1 and p170.2.
XX
XX DE C4bp; gp120; HIV; T lymphocyte; fusion protein.
XX
XX OS Homo sapiens.
XX
XX FH Key
XX FT Peptide 1..25
XX FT /label= signal_peptide
XX FT 25..132
XX FT /label= Ig-related
XX FT /note= "extracellular"
XX FT 41..109
XX FT Disulfide-bond
XX FT 133..202
XX FT /label= Ig-related
XX FT /note= "extracellular"
XX FT 155..184
XX FT Disulfide-bond
XX FT 203..318
XX FT /label= Ig-related
XX FT /note= "extracellular"
XX FT 319..395
XX FT Domain
XX FT /label= Ig-related
XX FT /note= "extracellular"
XX FT 328..370
XX FT Disulfide-bond
XX FT 356..416
XX FT Region
XX FT /label= transmembrane
XX FT 417..456
XX FT /label= cytoplasmic
XX
XX PN WO9111461-A.
XX
XX PD 08-AUG-1991.
XX
XX PF 26-JAN-1990; 90US-00470888.
XX
XX PR 26-JAN-1990; 90US-00470888.
XX
XX PA (BIOU ) BIOGEN INC.
XX
XX PI Pasek MP, Winkler G, Liu TR;
XX
XX DR WPI, 1991-252613/34.
XX
XX DR N-PSDB; AAQ13243.
XX
XX PT New C4 binding protein fusion proteins and DNA encoding them - comprise
XX PT assemblies of C4bp monomers linked to functional moiety, e.g. AZT, useful
XX PT as delivery vehicles in diagnosis and therapy.
XX
XX PS Example 3; Fig 3; 105pp; English.
XX
XX CC This is the preferred CD4 sequence for use in the construction of fusion
XX CC proteins with C4-binding protein. Truncated, soluble versions of CD4 can
XX CC also be used. The C4bp-CD4 fusion protein may be useful to target AZT or
XX CC similar anti-retroviral agent to HIV-infected cells. See AAQ13242-51.
XX CC (Updated on 25-MAR-2003 to correct PA field.)
XX
XX SQ Sequence 458 AA;

Query Match 59.0%; Score 2013; DB 2; Length 458;
Best Local Similarity 99.5%; Pred. No. 8.8e-101;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTQLALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 60
DB 62 MNRGVPRHLLVLTQLALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHKNSNOIK 121
QY 61 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLKIENSDTYICEVEDQKEVOL 120
DB 122 ILGNQGSFLTKGSPSKLNDRAISRSLMDQGNFPLIINKLKIENSDTYICEVEDQKEVOL 181
QY 181 TWICTTVLQONKVFVKIDIVVLAFOKASSIVYKKEGVQVFPPLATTVTEKLTGSGELMW 240
DB 242 TWICTTVLQONKVFVKIDIVVLAFOKASSIVYKKEGVQVFPPLATTVTEKLTGSGELMW 301
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTPDPLQMGKTLPHLTLTPOALPOYASSGNLTLLA 300
DB 302 QAERASSSKSWITFDLKNKEVSVKRVTPDPLQMGKTLPHLTLTPOALPOYASSGNLTLLA 361
QY 301 LEAKTGLHQBVLVVMRATQLOKNTLCEVWGPTSPKLMLSLKLENKAVSKREKRVW 360
DB 362 LEAKTGLHQBVLVVMRATQLOKNTLCEVWGPTSPKLMLSLKLENKAVSKREKRVW 421
QY 361 LNPAGMWOCILSDSGVLLSNIKVLPWTSTPV 364
DB 422 LNPAGMWOCILSDSGVLLSNIKVLPWTSTPV 455

```



QY 121 LVFGLTANSPTHLTQGGSLTTLTLESPPGSSPSVOCSPRGKNIQGGKTLVSQLELQDSG 180  
 DB 121 LVFGLTANSPTHLTQGGSLTTLTLESPPGSSPSVOCSPRGKNIQGGKTLVSQLELQDSG 180  
 QY 181 TWCTCTVLONQKQVEFKIDIVLAFQKASSIYKKKEGEQVEFSPPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTCTVLONQKQVEFKIDIVLAFQKASSIYKKKEGEQVEFSPPLAFTVEKLTGSGELMW 240  
 QY 241 QAERASSKSWITFDLKNKEVSVRVYODPKLQMGKKLPLHLTLPOLPOYASGSGNLTLA 300  
 DB 241 QAERASSKSWITFDLKNKEVSVRVYODPKLQMGKKLPLHLTLPOLPOYASGSGNLTLA 300  
 QY 301 LEAKTGKLAHQEVNLVVRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPEVMV 360  
 DB 301 LEAKTGKLAHQEVNLVVRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPEVMV 360  
 QY 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPV 394  
 DB 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPV 394

## RESULT 71

AAI39825  
 ID AAY39825 standard; protein; 394 AA.

XX AAY39825;

DT 03-DEC-1999 (first entry)

DE Soluble human T4 protein.

KM Soluble T4 protein; sT4; human; HIV; binding inhibitor; T4+ cell; AIDS;

KW vaccine; immunisation; therapy.

OS Homo sapiens.

PN US958678-A.

XX 28-SEP-1999.

PF 12-DEC-1994; 94US-00354452.

XX 21-AUG-1986; 86US-00898587.

PR 11-JUN-1991; 91US-00713564.

PR 06-JUL-1992; 92US-00909021.

XX (UYCO ) UNIV COLUMBIA NEW YORK.

PA (UYCO ) UNIV COLUMBIA NEW YORK.

PI McDougal JS, Weiss R, Axel R, Littman DR, Maddon PJ, Chess L;

XX WPI; 1999-561025/47.

DR N-PSDB; AA220694.

XX Human T4 protein inhibits HIV binding to T4 cells, useful for treating

PT AIDS.

XX Human T4 protein inhibits HIV binding to T4 cells, useful for treating

PS Disclosure; Col 13-16; 58pp; English.

XX This sequence represents the soluble human T4 protein of the invention.

CC The soluble human T4 protein blocks the binding of HIV to T4+ cells and

CC is therefore useful for the treatment of AIDS. Monoclonal antibodies

CC against the T4 protein may be used as vaccines for immunising subjects

XX against AIDS

XX Sequence 394 AA;

Query Match 58.9%; Score 2012; DB 2; Length 394;

Best Local Similarity 99.2%; Pred. No. 8.5e-101;

Matches 391; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRCVPRHLLLVQLALLPAAATGGRVVLGKKGDVETLCTASQKSIQFHMKNNOIK 60

DB 1 MNRCVPRHLLLVQLALLPAAATGGRVVLGKKGDVETLCTASQKSIQFHMKNNOIK 60

QY 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQNPFLITKNLKIENDSPYICEVEQKEEVOL 120  
 DB 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQNPFLITKNLKIENDSPYICEVEQKEEVOL 120  
 QY 121 LVFGLTANSPTHLTQGGSLTTLTLESPPGSSPSVOCSPRGKNIQGGKTLVSQLELQDSG 180  
 DB 121 LVFGLTANSPTHLTQGGSLTTLTLESPPGSSPSVOCSPRGKNIQGGKTLVSQLELQDSG 180  
 QY 181 TWCTCTVLONQKQVEFKIDIVLAFQKASSIYKKKEGEQVEFSPPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTCTVLONQKQVEFKIDIVLAFQKASSIYKKKEGEQVEFSPPLAFTVEKLTGSGELMW 240  
 QY 241 QAERASSKSWITFDLKNKEVSVRVYODPKLQMGKKLPLHLTLPOLPOYASGSGNLTLA 300  
 DB 241 QAERASSKSWITFDLKNKEVSVRVYODPKLQMGKKLPLHLTLPOLPOYASGSGNLTLA 300  
 QY 301 LEAKTGKLAHQEVNLVVRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPEVMV 360  
 DB 301 LEAKTGKLAHQEVNLVVRATOLQKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPEVMV 360  
 QY 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPV 394  
 DB 361 LNPEAGMWQCLLSDSGQVLLSNIKVLPTWSTPV 394

## RESULT 72

AAI39087  
 ID AAG79087 standard; protein; 458 AA.

XX AAG79087;

DT 10-DEC-2001 (first entry)

DE Amino acid sequence of human CD4 protein.

KM Human; receptor; DC-SIGN; dendritic cell; T lymphocyte; HIV; gp120;

KW C-type lectin; ICM3; HIV entry; T cell; macrophage; HIV infection; CD4.

OS Homo sapiens.

PN WO200164752-A2.

XX 07-SEP-2001.

XX 28-FEB-2001; 2001WO-US006322.

PR 02-MAR-2000; 2000US-00517605.

XX (UTNY ) UNIV NEW YORK STATE.

PA (UTNY ) UNIV NEW YORK STATE.

PI Littman DR, Kwon D, Van Kooyk Y, Geijtenbeek T;

XX WPI; 2001-602565/68.

DR An antibody for the treatment or prevention of HIV-infection comprises a

PT gp120 portion which binds to DC-SIGN or is exposed upon gp120 binding of

PT DC-SIGN due to concomitant conformational change.

XX Disclosure; Page 115-116; 131pp; English.

XX The specification describes an antibody which is specific for an

CC antigenic fragment of gp120. This antigenic fragment binds to DC-SIGN or

CC is exposed upon gp120 binding of DC-SIGN due to concomitant

CC conformational change. DC-SIGN is a receptor that is specifically

CC expressed on dendritic cells and facilitates infection of T lymphocytes

CC with HIV. DC-SIGN is identical to a HIV-1 gp120-binding C-type lectin. DC

CC -SIGN binds ICM-3 (which is expressed constitutively on T lymphocytes)

CC with high affinity. The antibody of the invention inhibits the trans

CC enhancement of HIV entry into a T cell or macrophage facilitated by

CC dendritic cells. The antibody is useful to treat or prevent HIV

CC infection. The present sequence represents a human CD4 protein

XX  
SQ Sequence 458 AA; 58.8%; Score 2007; DB 4; Length 458;  
Query Match Best Local Similarity 99.2%; Pred. No. 1.8e-100;  
Matches 391; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLVQLALIPATQGNKVVLGKKGDTVELTCTASQKKSIOFHWKNSNQIK 60  
DB 1 MNRGVPFRHLVLVQLALIPATQGNKVVLGKKGDTVELTCTASQKKSIOFHWKNSNQIK 60  
QY 61 ILGNQSSFLTKGPSKLNDRADSRSLMDQGNPPLIIKNIKIEDSDTYICEVEDQKEEYVL 120  
DB 61 ILGNQSSFLTKGPSKLNDRADSRSLMDQGNPPLIIKNIKIEDSDTYICEVEDQKEEYVL 120  
QY 121 LVFGLTANSPTHLLOQOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLEIQDSG 180  
DB 121 LVFGLTANSPTHLLOQOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLEIQDSG 180  
QY 181 TWCTVLQNKQKVEFKIDIVLAFOKASSIVYKKEGEQVFSFPLAFVTEKLTGSGELMW 240  
DB 181 TWCTVLQNKQKVEFKIDIVLAFOKASSIVYKKEGEQVFSFPLAFVTEKLTGSGELMW 240  
QY 241 QAERASSSKSWITFDLKNKEVSVKRYTODPKLOMGKKLPLHLTLPALPOYAGSGNLTLA 300  
DB 241 QAERASSSKSWITFDLKNKEVSVKRYTODPKLOMGKKLPLHLTLPALPOYAGSGNLTLA 300  
QY 301 LEAKTGKLEHQBVLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKREKAVV 360  
DB 301 LEAKTGKLEHQBVLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKREKAVV 360  
QY 361 LNPEAGMWQCLSDSGQVLLBSNIKVLPWTSTPV 394  
DB 361 LNPEAGMWQCLSDSGQVLLBSNIKVLPWTSTPV 394

RESULT 73  
AAB07768 standard; protein; 394 AA.  
ID AAB07768;  
AC AAB07768;  
XX 07-NOV-2000 (first entry)  
XX DE The soluble extracellular domain of the T4 glycoprotein.  
XX KM Human; T4 glycoprotein; human immunodeficiency virus; HIV;  
XX KM envelope glycoprotein; AIDS; virus binding.  
XX OS Homo sapiens.  
XX PN US6093539-A.  
XX PD 25-JUL-2000.  
XX PF 06-JUN-1995; 95US-00466368.  
XX PR 21-AUG-1986; 86US-00898587.  
XX PR 11-JUN-1991; 91US-00713564.  
XX PR 06-JUL-1992; 92US-00909021.  
XX PR 12-DEC-1994; 94US-00354452.  
XX PA (UYCO ) UNIV COLUMBIA NEW YORK.  
XX PI Madden PJ, Chess L, Axel R, Weiss R, McDougal JS, Littman DR,  
XX DR WPI; 2000-505203/45.  
XX DR N-PSDB; AAA59351.  
XX PT New isolated nucleic acid encoding a human T cell surface protein and the  
XX PT soluble surface T4 glycoprotein that it encodes, useful as prophylaxis  
XX PT for treating a subject infected with human acquired immune deficiency  
XX PT syndrome virus.

XX  
PS Disclosure; Col 11-14; 69pp; English.  
XX CC The present sequence represents an aqueous-soluble polypeptide comprising  
XX CC a portion of a human T4 glycoprotein. The portion specifically forms a  
XX CC complex with a human immunodeficiency virus (HIV) envelope glycoprotein.  
XX CC The DNA is useful for producing the soluble surface T4 glycoprotein. The  
XX CC soluble surface T4 glycoprotein is useful as a therapeutic agent, i.e. as  
XX CC prophylaxis for treating a subject infected with an HIV virus. Thus, the  
XX CC soluble T4 glycoprotein is useful for treating human AIDS. The soluble T4  
XX CC glycoprotein is also useful in diagnostic or screening assays, e.g. for  
XX CC screening inhibitors of virus binding, or for detecting and quantitating  
XX CC T4, T4+ cells and antibodies to T4, which are of diagnostic value for  
XX CC AIDS

SQ Sequence 394 AA; 58.6%; Score 2001; DB 3; Length 394;  
Query Match Best Local Similarity 98.7%; Pred. No. 3.3e-100;  
Matches 389; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLVQLALIPATQGNKVVLGKKGDTVELTCTASQKKSIOFHWKNSNQIK 60  
DB 1 MNRGVPFRHLVLVQLALIPATQGNKVVLGKKGDTVELTCTASQKKSIOFHWKNSNQIK 60  
QY 61 ILGNQSSFLTKGPSKLNDRADSRSLMDQGNPPLIIKNIKIEDSDTYICEVEDQKEEYVL 120  
DB 61 ILGNQSSFLTKGPSKLNDRADSRSLMDQGNPPLIIKNIKIEDSDTYICEVEDQKEEYVL 120  
QY 121 LVFGLTANSPTHLLOQOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLEIQDSG 180  
DB 121 LVFGLTANSPTHLLOQOSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLEIQDSG 180  
QY 181 TWCTVLQNKQKVEFKIDIVLAFOKASSIVYKKEGEQVFSFPLAFVTEKLTGSGELMW 240  
DB 181 TWCTVLQNKQKVEFKIDIVLAFOKASSIVYKKEGEQVFSFPLAFVTEKLTGSGELMW 240  
QY 241 QAERASSSKSWITFDLKNKEVSVKRYTODPKLOMGKKLPLHLTLPALPOYAGSGNLTLA 300  
DB 241 QAERASSSKSWITFDLKNKEVSVKRYTODPKLOMGKKLPLHLTLPALPOYAGSGNLTLA 300  
QY 301 LEAKTGKLEHQBVLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKREKAVV 360  
DB 301 LEAKTGKLEHQBVLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKAKVSKREKAVV 360  
QY 361 LNPEAGMWQCLSDSGQVLLBSNIKVLPWTSTPV 394  
DB 361 LNPEAGMWQCLSDSGQVLLBSNIKVLPWTSTPV 394

RESULT 74  
ADE65841 standard; protein; 458 AA.  
ID ADE65841;  
AC ADE65841;  
XX 29-JUN-2004 (first entry)  
XX DE Human CD4 receptor.  
XX DE Human CD4 receptor.  
XX KM Human; CD4 receptor; receptor; protein-protein interaction;  
XX KM protein array; PDZ domain; drug target screening.  
XX OS Homo sapiens.  
XX PN US2003170723-A1.  
XX PD 11-SEP-2003.  
XX PR 06-MAR-2002; 2002US-00092138.  
XX PR 06-MAR-2002; 2002US-00092138.  
XX PR 06-MAR-2002; 2002US-00092138.

PA (SATO/) SATO T.  
 XX  
 XX  
 PI Sato T;  
 XX  
 XX WPI: 2003-852032/79.  
 XX  
 XX Preparing a protein array useful for screening drug targets comprises  
 PT depositing an array of a first protein on substrate, and applying a  
 PT second protein comprising an amino acid sequence that binds to a domain  
 PT of the first protein.  
 XX  
 XX Disclosure; SEQ ID NO 25; 60pp; English.  
 PS  
 XX  
 CC The invention relates to a method for preparing a protein array based on  
 CC protein-protein interaction, by depositing an array of a first protein  
 CC comprising a PDZ domain on a substrate, and applying a second protein  
 CC comprising an amino acid sequence that binds to the PDZ domain of the  
 CC first protein. The method is useful for preparing protein arrays based on  
 CC biochemical protein-protein interactions. Arrays produced by this method  
 CC are useful for screening drug targets. This sequence represents the human  
 CC CD4 receptor, used in the method of the invention.  
 CC  
 XX  
 SQ Sequence 458 AA;  
 Query Match 58.5%; Score 1997; DB 7; Length 458;  
 Best Local Similarity 98.7%; Pred. No. 6.4e-100;  
 Matches 389; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MNRGVPRHLLVLTQALLPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60  
 DB 1 MNRGVPRHLLVLTQALLPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60  
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIKNKIKIEDSDTYICEVEDQKEEVOL 120  
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIKNKIKIEDSDTYICEVEDQKEEVOL 120  
 QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWCTCTVLQNGKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTCTVLQNGKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
 QY 241 QAERASSSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300  
 DB 241 QAERASSSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300  
 QY 301 LEATGKGLHQEVNLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPYMV 360  
 DB 301 LEATGKGLHQEVNLVVMRATOLQKNTLCEVWGPTSPKMLSLKLENKEAKVSKREKPYMV 360  
 QY 361 LNPAGMWQCLLSDSGQVLTLESNIKVLPTWSTPV 394  
 DB 361 LNPAGMWQCLLSDSGQVLTLESNIKVLPTWSTPV 394  
 RESULT 75  
 AAR06373  
 ID AAR06373 standard; protein; 458 AA.  
 XX  
 AC AAR06373;  
 XX  
 DT 31-OCT-2002 (revised)  
 DT 20-DEC-1990 (first entry)  
 XX  
 XX T4 encoded by plasmid p170-2.  
 DE  
 XX plasmid p170-2; soluble T4 protein; AIDS; ARC; HIV.  
 KM  
 XX Synthetic.  
 OS  
 XX Key Location/Qualifiers

FT Peptide  
 FT 1. 23  
 FT /label= hydrophobic/secretory signal  
 FT 24. 117  
 FT /label= extracellular  
 FT /note= "homology to V-regions"  
 FT 118. 132  
 FT /label= extracellular  
 FT /note= "homology to J-regions"  
 FT 133. 397  
 FT /label= extracellular  
 FT /note= "glycosylated region"  
 FT 398. 418  
 FT /label= transmembrane sequence  
 FT /note= "hydrophobic"  
 FT 419. 458  
 FT /label= intracytoplasmic  
 FT /note= "very hydrophilic"  
 PN W09008198-A.  
 XX 26-JUL-1990.  
 PD  
 XX 18-JAN-1989; 89US-00300096.  
 PF  
 XX 18-JAN-1989; 89US-00300096.  
 PR  
 XX (HARD ) HARVARD COLLEGE.  
 PA  
 XX Letvin NA;  
 PI  
 XX WPI: 1990-254040/33.  
 DR N-PSDB; AAQ05607.  
 XX  
 PT Treating or preventing AIDS, ARC or HIV infection - by administering an  
 PT immunologically effective amt. of soluble T4 protein.  
 XX  
 PS Disclosure; Fig 1; 121pp; English.  
 CC Soluble T4 can be produced by truncating the CDS to remove the  
 CC transmembrane and cytoplasmic domains. The soluble forms may be modified  
 CC to increase their immunogenicity by addition of an adjuvant such as  
 CC incomplete Freund's adjuvant. The T4 interferes with HIV/T4 interaction  
 CC and elicits anti-soluble T4 antibody production. Soluble T4 proteins  
 CC include the following polypeptide fragments: amino acids 1-385, 24-385,  
 CC Met-24-385, 24-397, 1-400 and Met-24-400. See also AAQ05606. (Updated on  
 CC 31-OCT-2002 to add missing OS field.)  
 CC  
 XX  
 SQ Sequence 458 AA;  
 Query Match 58.4%; Score 1993; DB 2; Length 458;  
 Best Local Similarity 99.0%; Pred. No. 1e-99;  
 Matches 390; Conservative 0; Mismatches 4; Indels 0; Gaps 0;  
 QY 1 MNRGVPRHLLVLTQALLPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60  
 DB 1 MNRGVPRHLLVLTQALLPAATQGNKVLGKGGDTVELTCTASQKKSIOFHMKNNSQIK 60  
 QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIKNKIKIEDSDTYICEVEDQKEEVOL 120  
 DB 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFPLIKNKIKIEDSDTYICEVEDQKEEVOL 120  
 QY 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 DB 121 LVFGLTANSDTHLLQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
 QY 181 TWCTCTVLQNGKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTCTVLQNGKVEFKIDIVLAFQKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
 QY 241 QAERASSSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300  
 DB 241 QAERASSSSKSWITFDLKNKEVSVKRVYODPKLQMGKKLPHLTLPLQALPOYAGSGNLTIA 300

```

QY 301 LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVV 360
DB 301 LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVV 360
QY 361 LNPEAGMWOCLLSDSGOVLLESNIKVLPTWSTPV 394
DB 361 LNPEAGMWOCLLSDSGOVLLESNIKVLPTWSTPV 394

RESULT 76
AAP94703
ID AAP94703 standard; protein; 524 AA.
XX
AC AAP94703;
XX
DT 25-MAR-2003 (revised)
DT 22-MAR-1991 (first entry)
XX
DE Sequence encoded by T4 lymphocyte cDNA obtained from PBL clone lambda-203
-4.
XX
KM HIV; soluble T4; immunotherapeutic; prophylactic; diagnostic; AIDS; ARC.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 67 /note= "AA DESIGNATED NUMBER -23"
FT Misc-difference 90 /note= "AA DESIGNATED NUMBER 1"
FT Misc-difference 92 /note= "MATURE N-TERMINUS"
FT
FT
XX
XX WO8901940-A.
XX
XX 09-MAR-1989.
XX
XX 01-SEP-1988; 88MO-US002940.
XX
XX 04-SEP-1987; 87US-00094322.
XX 07-JAN-1988; 88US-00141649.
XX
XX (BIOJ ) BIOGEN INC.
XX
XX Fisher RA, Gilbert W, Sato VL, Flavell RA, Maganore JM;
XX
XX WPI; 1989-085519/11.
XX
XX N-PSDB; AAN90642.
XX
XX
XX DNA sequences coding for soluble T4-like polypeptide(s) - used in
XX immunotherapeutic and immunosuppressive compns. and for preventing,
XX treating or detecting AIDS.
XX
XX Disclosure; Fig 3; 207pp; English.
XX
XX The polypeptides encoded are useful in immunotherapeutic, prophylactic
XX and diagnostic compns. They can be used to purify HIV from a sample. The
XX soluble T4 protein-based compns. are useful in treating immunodeficient
XX patients suffering from diseases caused by agents whose primary targets
XX are T4+ lymphocytes. They can be used for preventing, treating or
XX detecting AIDS, ARC and HIV infection. (Updated on 25-MAR-2003 to correct
XX PR field.)
XX
XX
XX Sequence 524 AA;
XX
XX Query Match 58.4%; Score 1993; DB 1; Length 524;
XX Best Local Similarity 99.0%; Pred. No. 1.2e-99;
XX Matches 390; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 1 MNRGVPFRHLILVQLALLPAPATGKNVLAGKGDVETLCTASQKKSIQFHWKNSNQIK 60
DB 67 MNRGVPFRHLILVQLALLPAPATGKNVLAGKGDVETLCTASQKKSIQFHWKNSNQIK 126

```

```

QY 61 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVDQKEEYQL 120
DB 127 ILGNQGSFLTKGSPSKLNDRADSRSLMDQGNFLLIKNLKIEDSDTYICEVDQKEEYQL 186
QY 121 LVFGLTANSDBTHLQGGSLITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180
DB 187 LVFGLTANSDBTHLQGGSLITLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 246
QY 181 TWTCTVLQNKQKVEERKIDIVLAFOKASSIYVKKEGEVFEFPLATVTEKLTSGGELMW 240
DB 247 TWTCTVLQNKQKVEERKIDIVLAFOKASSIYVKKEGEVFEFPLATVTEKLTSGGELMW 306
QY 241 QAEARASSKSWITTPDLKKNKVEVKRVTQDPKIQMGKPLHLTLPLQALPQVAGSGULTLA 300
DB 307 QAEARASSKSWITTPDLKKNKVEVKRVTQDPKIQMGKPLHLTLPLQALPQVAGSGULTLA 366
QY 301 LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVV 360
DB 367 LEAKTGKLGHOEVNLYVMRATOLQKNLTCEVWGPTSPKMLSLKLENKAVSKREKPVV 426
QY 361 LNPEAGMWOCLLSDSGOVLLESNIKVLPTWSTPV 394
DB 427 LNPEAGMWOCLLSDSGOVLLESNIKVLPTWSTPV 460

RESULT 77
AAR07640
ID AAR07640 standard; protein; 2458 AA.
XX
XX AAR07640;
XX
XX 31-OCT-2002 (revised)
XX 20-DEC-1990 (first entry)
XX
XX Deduced protein sequence of p170-2 comprising T4.
XX
XX plasmid p170-2; soluble T4 protein; AIDS; ARC; HIV.
XX
XX Synthetic.
XX
XX Key Location/Qualifiers
XX FH Protein 400..858
XX FT /label= T4 surface glycoprotein
XX
XX
XX PN MO908198-A.
XX
XX 26-JUL-1990.
XX
XX 18-JAN-1989; 89US-00300096.
XX
XX 18-JAN-1989; 89US-00300096.
XX
XX (HARD ) HARVARD COLLEGE.
XX
XX Letvin NA;
XX
XX WPI; 1990-254040/33.
XX N-PSDB; AAQ05607.
XX
XX Treating or preventing AIDS, ARC or HIV infection - by administering an
XX immunologically effective amt. of soluble T4 protein.
XX
XX Disclosure; Fig 1; 121pp; English.
XX
XX Entire sequence from T4-encoding plasmid p170-2. It is almost identical
XX to the sequence published by Maddon et al.(1985) with the exception of
XX three codon changes. At T4 amino acid residue 3, (posn.403 of entire
XX sequence) Lys is encoded in stead of Asn. At posn. 64, (posn.464) Arg
XX replaces Trp and at posn. 231, (posn. 631) Ser replaces Phe. Soluble T4
XX can be produced by truncating the CDS to remove the transmembrane and
XX cytoplasmic domains. The soluble forms may be modified to increase their
XX immunogenicity by addition of an adjuvant such as incomplete Freund's
XX adjuvant. The T4 interferes with HIV/T4 interaction and elicits anti-

```

soluble T4 antibody production. See also AA005608. (Updated on 31-OCT-2002 to add missing OS field.)

Sequence 2458 AA:

Query Match 58.4%; Score 1993; DB 2; Length 2458;  
Best Local Similarity 99.0%; Pred. No. 6e-99;  
Matches 390; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVQLALPPAATQGNKVVLGKGDVETLTCTASQKKSIOFHMKNSQIK 60  
DB 400 MNRGVPRHLLVQLALPPAATQGNKVVLGKGDVETLTCTASQKKSIOFHMKNSQIK 459  
QY 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120  
DB 460 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 519  
QY 121 LVFGLTANSPTHLIQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
DB 520 LVFGLTANSPTHLIQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 579  
QY 181 TWCTVLOQNKKEVFKIDIVLAFOKASSIVYKKEGQVEFSFPLATFVEKLTGSGELMW 240  
DB 580 TWCTVLOQNKKEVFKIDIVLAFOKASSIVYKKEGQVEFSFPLATFVEKLTGSGELMW 639  
QY 241 QAERASSSKSWITFDLNKKEVSRYKVTODPKLOMGKKLPHTLTPALPOYAGSGNLTLLA 300  
DB 640 QAERASSSKSWITFDLNKKEVSRYKVTODPKLOMGKKLPHTLTPALPOYAGSGNLTLLA 699  
QY 301 LEAKTGKLGHEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360  
DB 700 LEAKTGKLGHEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 759  
QY 361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 394  
DB 760 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 793

#### RESULT 78

AA04031 standard; protein; 2458 AA.

AC AAR04031;  
XX  
DT 25-MAR-2003 (revised)  
DT 31-OCT-2002 (revised)  
DT 29-MAY-1990 (first entry)  
XX  
DE Full length T4 encoded by plasmid p170-2.  
XX  
KM Soluble T4; p170-2; anti-retroviral agent; AIDS; ARC; HIV; AZT.  
XX  
OS Synthetic.  
XX  
XX Key Location/Qualifiers  
FH M18c-difference 423  
FT M18c-difference 425 /note= "TTP of Madden et al replaced by Arg"  
FT M18c-difference 425 /note= "Asp of Madden et al replaced by Lys"  
FT M18c-difference 653 /note= "Phe of Madden et al replaced by Ser"  
XX  
XX MO8911860-A.  
XX  
XX 14-DEC-1989.  
XX  
XX 08-JUN-1989; 89MO-US002453.  
XX  
XX 10-JUN-1988; 88US-00204645.  
XX 20-APR-1989; 89US-00341080.  
XX  
XX (BIOJ ) BIOGEN NV INC.  
XX (GEHO ) GEN HOSPITAL CORP.

PA (BIOJ ) BIOGEN INC.  
PA (BIOJ ) BIOGEN INC.  
XX  
XX

PI Fisher RA, Schooley RT, Hirsch MS, Johnson VA, Walker BD;  
XX

DR WPI; 1990-007302/01.  
XX

DR N-PSDB; AA003005.  
XX

PT Combinations of soluble T4 protein and anti-retroviral agent - having synergistic activity in treatment and prevention of AIDS, arc and HIV infection.  
PT  
XX

PS Disclosure; Fig 1; 100pp; English.  
XX

CC The sequence differs from that determined by PJ Madden et al., [Cell, 42 pp. 93-104 (1985)] in three places due to three nucleotide substitutions. CC  
CC The Asp reported at position 3 by Madden et al. was the result of a CC  
CC sequencing error [DR Litzman et al. Cell, 55, p.541 (1988)]. X = stop CC  
CC codon. The sequence was deduced from the cDNA insert of p170-2. Soluble CC  
CC T4 constructs may be produced by truncating this sequence to give CC  
CC fragments from position 400 to 799, removing the transmembrane and CC  
CC intracytoplasmic domains whilst retaining the extracellular region CC  
CC responsible for HIV binding. The sol. T4 is combined with an anti-viral CC  
CC agent such as AZT. See also AA003006. (Updated on 31-OCT-2002 to add CC  
CC missing OS field.) (Updated on 25-MAR-2003 to correct PA field.)  
XX

SO Sequence 2458 AA:

Query Match 58.4%; Score 1993; DB 2; Length 2458;  
Best Local Similarity 99.0%; Pred. No. 6e-99;  
Matches 390; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVQLALPPAATQGNKVVLGKGDVETLTCTASQKKSIOFHMKNSQIK 60  
DB 400 MNRGVPRHLLVQLALPPAATQGNKVVLGKGDVETLTCTASQKKSIOFHMKNSQIK 459  
QY 61 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 120  
DB 460 ILGNQGSFLTKGSPKLNDRADSRSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVOL 519  
QY 121 LVFGLTANSPTHLIQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 180  
DB 520 LVFGLTANSPTHLIQGOSLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDQSG 579  
QY 181 TWCTVLOQNKKEVFKIDIVLAFOKASSIVYKKEGQVEFSFPLATFVEKLTGSGELMW 240  
DB 580 TWCTVLOQNKKEVFKIDIVLAFOKASSIVYKKEGQVEFSFPLATFVEKLTGSGELMW 639  
QY 241 QAERASSSKSWITFDLNKKEVSRYKVTODPKLOMGKKLPHTLTPALPOYAGSGNLTLLA 300  
DB 640 QAERASSSKSWITFDLNKKEVSRYKVTODPKLOMGKKLPHTLTPALPOYAGSGNLTLLA 699  
QY 301 LEAKTGKLGHEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 360  
DB 700 LEAKTGKLGHEVNLVVMRATQLOKNLTCEVWGPTSPKMLSLKLENKEAKVSKREKPVW 759  
QY 361 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 394  
DB 760 LNPEAGMWQCLSDSGVLLSNIKVLPTWSTPV 793

#### RESULT 79

AA020151 standard; protein; 399 AA.

AC AAR20151;  
XX  
XX  
DT 25-MAR-2003 (revised)  
DT 31-MAR-1992 (first entry)  
XX  
XX Chimpanzee sol. CD4 encoded by pSQ200.  
XX  
XX Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;

acquired immune deficiency syndrome; AIDS related complex;  
T helper lymphocytes.  
Pan troglodytes.  
Key Location/Qualifiers  
FT Peptide 1..25  
/label= signal\_sequence  
WO9118618-A.  
XX  
XX 12-DEC-1991.  
XX  
XX PD  
XX PF 25-MAY-1990; 90US-00529186.  
XX  
XX PR 25-MAY-1990; 90US-00529186.  
XX  
XX PA (BIOU ) BIOGEN INC.  
XX  
XX PI Fisher RA, Heseion C, Burkly LC;  
XX  
XX DR WPI; 1992-007200/01.  
XX  
XX DR N-PSDB; AAQ20326.  
XX  
XX PT New immuno-therapeutic human CD4 variants and derive. - elicit AB  
XX  
XX PT production to HIV gp.120, useful in treating, preventing and diagnosing  
XX  
XX PT AIDS, ARC and HIV infections.  
XX  
XX PS Claim 15; Fig 21; 179pp; English.  
XX  
XX CC The sequence was deduced from the DNA sequence of clone pSG200 which was  
XX  
XX CC obt'd. by cloning using a reverse transcriptase/PCR amplification  
XX  
XX CC procedure with a non-specific oligo dt to prime first strand synthesis.  
XX  
XX CC The PCR fragments were ligated into sequencing plasmid pNN08 (see  
XX  
XX CC AAQ20326) The DNA sequence can be used to express recombinant soluble CD4  
XX  
XX CC and analogues for use in diagnosis and treatment of diseases caused by  
XX  
XX CC infective agents whose primary targets are T4+ lymphocytes. See also  
XX  
XX CC AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA  
XX  
XX CC field.)  
XX  
XX SQ Sequence 399 AA;

Query Match 58.3%; Score 1991; DB 2; Length 399;  
Best Local Similarity 98.5%; Pred. No. 1.2e-99;  
Matches 388; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNOIK 60  
DB 1 MNRGVPRHLLVLTALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNOIK 60  
DB 1 MNRGVPRHLLVLTALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNOIK 60  
QY 61 ILGNQGSFLTKGPSKLNDRADRSRSLMDQGNFLLIINKLKIEDSDTYICEVGDQKEVOL 120  
DB 61 ILGNQGSFLTKGPSKLNDRADRSRSLMDQGNFLLIINKLKIEDSDTYICEVGDQKEVOL 120  
DB 121 LVFGLTANSDTHLLOGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180  
QY 121 LVFGLTANSDTHLLOGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180  
DB 121 LVFGLTANSDTHLLOGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180  
QY 181 TWTCVTLQONKVEFKIDIVLAFOKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
DB 181 TWTCVTLQONKVEFKIDIVLAFOKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
QY 241 QAEBASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPHLTLPOALPOYAGSGNLTLLA 300  
DB 241 QAEBASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPHLTLPOALPOYAGSGNLTLLA 300  
QY 301 LEAKTGLHOEVNLVNNRATQLOKNLTCEVWGPTSPKLMSTLKENKEAVSRKRPVWV 360  
DB 301 LEAKTGLHOEVNLVNNRATQLOKNLTCEVWGPTSPKLMSTLKENKEAVSRKRPVWV 360  
QY 361 LNPBAGMWQCLLSDSGOVLLESNIKVLPTWSTPV 394  
DB 361 LNPBAGMWQCLLSDSGOVLLESNIKVLPTWSTPV 394

RESULT 80  
AAR11285  
ID AAR11285 standard; protein; 458 AA.  
XX  
XX AAR11285;  
AC  
XX 25-MAR-2003 (revised)  
DT 09-JAN-2003 (revised)  
DT 29-APR-1991 (first entry)  
XX  
XX gp120 binding protein.  
XX  
XX Human; CD4; AIDS; HIV1; SIV; gp120.  
XX  
XX OS Unidentified.  
XX  
XX FH Key Location/Qualifiers  
FT Misc-difference 59  
FT /label= Thr or Ile  
FT FT  
FT Misc-difference 93  
FT /label= Thr or Pro  
XX  
XX PN EP414178-A.  
XX  
XX PD 27-FEB-1991.  
XX  
XX PF 18-AUG-1990; 90EP-00115877.  
XX  
XX PR 23-AUG-1989; 89US-00397782.  
XX  
XX PA (GHEO ) GEN HOSPITAL CORP.  
XX  
XX PI Seed B, Camerini D;  
XX  
XX DR WPI; 1991-059419/09.  
XX  
XX DR N-PSDB; AAQ10887.  
XX  
XX PT New non-human primate and human CD4 or gp120 molecules - used to treat  
XX  
XX PT HIV or SIV and immunoglobulin and gp120 binding molecules from new fusion  
XX  
XX PT proteins.  
XX  
XX PS Claim 17; Page 57; 87pp; English.  
XX  
XX CC The fragment from residues 1-134 is also independently claimed. The sub-  
XX  
XX CC fragment (and the complete polypeptide) can bind to HIV gp120. See also  
XX  
XX CC AAQ10885-6, AAQ10888. (Updated on 09-JAN-2003 to add missing OS field.)  
XX  
XX CC (Updated on 25-MAR-2003 to correct PA field.)  
XX  
XX SQ Sequence 458 AA;

Query Match 58.3%; Score 1991; DB 2; Length 458;  
Best Local Similarity 98.5%; Pred. No. 1.3e-99;  
Matches 388; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLVLTALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNOIK 60  
DB 1 MNRGVPRHLLVLTALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNOIK 60  
DB 1 MNRGVPRHLLVLTALPAATQGNKVVLGKKGDTVELTCTASQKKSIOFHMKNNOIK 60  
QY 61 ILGNQGSFLTKGPSKLNDRADRSRSLMDQGNFLLIINKLKIEDSDTYICEVGDQKEVOL 120  
DB 61 ILGNQGSFLTKGPSKLNDRADRSRSLMDQGNFLLIINKLKIEDSDTYICEVGDQKEVOL 120  
DB 121 LVFGLTANSDTHLLOGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180  
QY 121 LVFGLTANSDTHLLOGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180  
DB 121 LVFGLTANSDTHLLOGSLTLTLSPGSSPSVQCRSPRGKNIQGGKTLVSQLELODSG 180  
QY 181 TWTCVTLQONKVEFKIDIVLAFOKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
DB 181 TWTCVTLQONKVEFKIDIVLAFOKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
QY 241 QAEBASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPHLTLPOALPOYAGSGNLTLLA 300

```

Db      241 EAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPHLTLPQALPYAGSGNLTLA 300
Qy      301 LEAKTGKHQEVNLVVMRAATOLQKNTCEVWGPTSPKMLSLKLENKAKVSKREKPYWV 360
Db      301 LEAKTGKHQEVNLVVMRAATOLQKNTCEVWGPTSPKMLSLKLENKAKVSKREKAVW 360
Qy      361 LNPEAGMWQCLSDSGQVLLSNIKVLPTWSTPV 394
Db      361 LNPEAGMWQCLSDSGQVLLSNIKVLPTWSTPV 394

```

## RESULT 81

```

AAR10988 ID AAR10988 standard; protein; 458 AA.
XX
XX AAR10988;
AC
XX 25-MAR-2003 (revised)
DT 29-APR-1991 (first entry)
XX
XX Chimpanzee CD4 protein.
DE
XX Chimpanzee; CD4; AIDS; HIV1; SIV.
KM
XX Pan troglodytes.
OS
XX
XX Key Location/Qualifiers
FH 26
FT Protein /label= mature CD4
XX
XX EP414178-A.
XX
XX 27-FEB-1991.
PD
XX 18-AUG-1990; 90EP-00115877.
PF
XX 23-AUG-1989; 89US-00397782.
PR
XX
XX (GENO ) GEN HOSPITAL CORP.
PA
XX Seed B, Camerini D;
PI
XX WPI; 1991-059419/09.
XX N-PSDB; AAQ10886.
DR
XX
XX New non-human primate and human CD4 or gp120 molecules - used to treat
PT HIV or SIV and immunoglobulin and gp120 binding molecules from new fusion
PT proteins.
XX
XX Claim 4; Page 45; 87pp; English.
PS
XX
XX The CD4 protein or HIV gp120-binding fragments of it are used to detect
CC and treat HIV and SIV infection. Animals which can be treated include
CC humans, baboons, orang-utans, chimpanzees, gorillas and rhesus monkey.
CC The chimpanzee CD4 is 99 per cent homologous to its human counterpart,
CC possessing 5 amino acid substitutions in the 433 residue predicted mature
CC polypeptide. See also AAQ10885, AAQ10887-8. (Updated on 25-MAR-2003 to
CC correct PA field.)
XX
XX Sequence 458 AA;
SQ

```

## Query Match

58.3%; Score 1991; DB 2; Length 458;  
 Bect Local Similarity 98.5%; Pred. No. 1.3e-99;  
 Matches 388; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

Qy      1 MNRGVPSRHLVLTQALLPATQGNKVVLGKGGTVELTCTASOKSIOFHMKNQIK 60
Db      1 MNRGVPSRHLVLTQALLPATQGNKVVLGKGGTVELTCTASOKSIOFHMKNQIK 60
Qy      61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFIINKLKIETSDTYICEVEDQKEEVOL 120
Db      61 ILGNQGSFLTKGPSKLNDRADSRSLMDQGNPFIINKLKIETSDTYICEVEDQKEEVOL 120

```

```

Qy      121 LVFGLTANSPDTHLLQGSLLTLESPPGSSPSVOCRSRPRGNIOGKTLSSVSOLELQDSG 180
Db      121 LVFGLTANSPDTHLLQGSLLTLESPPGSSPSVOCRSRPRGNIOGKTLSSVSOLELQDSG 180
Qy      181 TWCTVTLQNKQKVEFKIDIVLAFQKASSIVYKKEGQVEFSFPLAFTVEKLTGSGELMW 240
Db      181 TWCTVTLQNKQKVEFKIDIVLAFQKASSIVYKKEGQVEFSFPLAFTVEKLTGSGELMW 240
Qy      241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPHLTLPQALPYAGSGNLTLA 300
Db      241 QAERASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLPHLTLPQALPYAGSGNLTLA 300
Qy      301 LEAKTGKHQEVNLVVMRAATOLQKNTCEVWGPTSPKMLSLKLENKAKVSKREKPYWV 360
Db      301 LEAKTGKHQEVNLVVMRAATOLQKNTCEVWGPTSPKMLSLKLENKAKVSKREKAVW 360
Qy      361 LNPEAGMWQCLSDSGQVLLSNIKVLPTWSTPV 394
Db      361 LNPEAGMWQCLSDSGQVLLSNIKVLPTWSTPV 394

```

## RESULT 82

```

AAR20150 ID AAR20150 standard; protein; 400 AA.
XX
XX AAR20150;
AC
XX 25-MAR-2003 (revised)
DT 31-MAR-1992 (first entry)
XX
XX Chimpanzee sol. CD4 sequence from psQ205.
DE
XX
XX Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;
KM acquired immune deficiency syndrome; AIDS related complex;
KM T helper lymphocytes.
XX
XX Pan troglodytes.
OS
XX
XX Key Location/Qualifiers
FH 1.25
FT Peptide /label= signal_sequence
XX
XX W09118618-A.
XX
XX 12-DEC-1991.
PD
XX 25-MAY-1990; 90US-00529186.
PF
XX 25-MAY-1990; 90US-00529186.
PR
XX 25-MAY-1990; 90US-00529186.
PA
XX (BIOU ) BIOGEN INC.
PA
XX Fisher RA, Hession C, Burkly LC;
PI
XX WPI; 1992-007200/01.
XX N-PSDB; AAQ20325.
DR
XX
XX New immuno-therapeutic human CD4 variants and deriva. - elicit AB
PT production to HIV gp.120, useful in treating, preventing and diagnosing
PT AIDS, ARC and HIV infections.
XX
XX Claim 15; Fig 20; 179pp; English.
PS
XX
XX The sequence was deduced from the DNA sequence of clone psQ205 which was
CC obd. by cloning using a reverse transcriptase/PCR amplification
CC procedure. The DNA sequence can be used to express recombinant soluble
CC CD4 and analogues for use in diagnosis and treatment of diseases caused
CC by infective agents whose primary targets are T4+ lymphocytes. See also
CC AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA
CC field.)
XX
XX Sequence 400 AA;
SQ

```





OS Homo sapiens.  
 XX EP325262-A.  
 XX 26-JUL-1989.  
 XX 20-JAN-1989; 89EP-00100913.  
 XX 22-JAN-1988; 88US-00147351.  
 XX (GCHO ) GEN HOSPITAL CORP.  
 XX Seed B;  
 XX WPI; 1989-214472/30.  
 XX N-PSDB; AAN90358.  
 DR Immunoglobulin-CD4 fusion proteins - used for treating HIV or SIV  
 PT Infection or detecting HIV or SIV in sample.  
 PS Example; Table 3, Page 34-40; 68pp; English.  
 XX The fusion protein genes of the invention pref. comprises cDNA sequences  
 CC which encode CD4 or a fragment which binds gp120 ligated to an expression  
 CC plasmid which encodes an antibody in which the variable region of the  
 CC gene has been deleted (see WO87-02671). The CD4 portion of the fusion  
 CC protein may comprise the complete CD4 sequence, the 370 AA extracellular  
 CC region and the membrane spanning domain, or the extracellular region. The  
 CC Ig heavy chain is pref. from 19M, 19G1 or 19G3. The following are  
 CC specifically claimed: fusion proteins CD4H1ambda1, CD4mu, CD4Pmu,  
 CC CD4E1ambda1, and CD4mu (No. 67608), CD4P1ambda (No. 67609) and  
 CC PCD4E1ambda1 (No. 67610). (Updated on 25-MAR-2003 to correct PA field.)  
 XX Sequence 399 AA;

Query Match 58.1%; Score 1982.5; DB 1; Length 399;  
 Best Local Similarity 94.9%; Pred. No. 3.4e-99;  
 Matches 391; Conservative 1; Mismatches 7; Indels 13; Gaps 1;

QY 1 MNRGVPFRHLLVQLALLPAATQGNKVLGKGDVETLCTASQKSIQFHMKNNOIK 60  
 DB 1 MNRGVPFRHLLVQLALLPAATQGNKVLGKGDVETLCTASQKSIQFHMKNNOIK 60  
 QY 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNPILITKLTIEDSDTYICEVEDQKEEYOL 120  
 DB 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNPILITKLTIEDSDTYICEVEDQKEEYOL 120  
 QY 121 LVFGLTANSPTHLLOGSLTLTLESPPGSSPSVOCRSRPRGNIOGKTLVSQLELODSG 180  
 DB 121 LVFGLTANSPTHLLOGSLTLTLESPPGSSPSVOCRSRPRGNIOGKTLVSQLELODSG 180  
 QY 121 LVFGLTANSPTHLLOGSLTLTLESPPGSSPSVOCRSRPRGNIOGKTLVSQLELODSG 180  
 DB 121 LVFGLTANSPTHLLOGSLTLTLESPPGSSPSVOCRSRPRGNIOGKTLVSQLELODSG 180  
 QY 181 TWCTCTVLONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTCTVLONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
 QY 241 QAERASSSKSWITPPLKNKEVSVKRVTDPLQNGKTLPLTLTPOALPOVAGSGNLTLA 300  
 DB 241 QAERASSSKSWITPPLKNKEVSVKRVTDPLQNGKTLPLTLTPOALPOVAGSGNLTLA 300  
 QY 301 LEATGKLGHOEVNLVVMKATOLQKLTCEVWGPTSPKMLSLKLENKEAKSKREKPYWV 360  
 DB 301 LEATGKLGHOEVNLVVMKATOLQKLTCEVWGPTSPKMLSLKLENKEAKSKREKPYWV 360  
 QY 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPVPCPAPEPKSCDKHTTCE 412  
 DB 361 LNPEAGMWQCLLSDSGVLLSNIKVLPTWSTPVPCPAPEPKSCDKHTTCE 412

RESULT 95  
 ID AAR20154 standard; protein; 400 AA.  
 XX AAR20154;  
 AC AAR20154;

XX 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 31-MAR-1992 (first entry)  
 XX Sol. rhesus-human chimeric CD4 encoded by pDG100.  
 DE  
 XX  
 KW Human immunodeficiency virus; HIV; gp 120; AIDS; ARC; glycoprotein;  
 KW acquired immune deficiency syndrome; AIDS related complex; monkey;  
 KW T helper lymphocytes.  
 XX  
 XX Macaca mulatta; (Rhesus).  
 OS Homo sapiens.  
 OS Chimeric.  
 XX  
 FH Key Location/Qualifiers  
 FT Region 1..131  
 FT /note= "rhesus CD4 encoded by pSQ146 (AAQ20328)"  
 FT Peptide 1..25  
 FT /label= signal\_sequence  
 FT Region 132..400  
 FT /note= "human CD4 encoded by PBG391 (US8802940)"

XX WO9118618-A.  
 XX 12-DEC-1991.  
 XX 25-MAY-1990; 90US-00529186.  
 XX 25-MAY-1990; 90US-00529186.  
 XX (BIOJ ) BIOGEN INC.  
 XX  
 XX Fisher RA, Hession C, Burkly LC;  
 XX WPI; 1992-007200/01.  
 DR N-PSDB; AAQ20329.  
 DR

PT New immuno-therapeutic human CD4 variants and derivs. - elicit AB  
 PT production to HIV gp.120, useful in treating, preventing and diagnosing  
 PT AIDS, ARC and HIV infections.  
 PS Claim 14; Fig 13; 179pp; English.

CC The sequence is encoded by pDG100, chimeric plasmid prepd. from human and  
 CC rhesus derived CD4 DNA. The plasmid can be used to express recombinant  
 CC sol. chimeric CD4 for use in diagnosis and treatment of diseases caused  
 CC by infective agents whose primary targets are T4+ lymphocytes. See also  
 CC AAR20148-R20155 and AAR21078. (Updated on 25-MAR-2003 to correct PA  
 CC field.) (Updated on 24-OCT-2003 to standardise OS field)

XX Sequence 400 AA;

Query Match 57.0%; Score 1946; DB 2; Length 400;  
 Best Local Similarity 95.7%; Pred. No. 3.1e-97;  
 Matches 377; Conservative 9; Mismatches 8; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLLVQLALLPAATQGNKVLGKGDVETLCTASQKSIQFHMKNNOIK 60  
 DB 1 MNRGVPFRHLLVQLALLPAATQGNKVLGKGDVETLCTASQKSIQFHMKNNOIK 60  
 QY 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNPILITKLTIEDSDTYICEVEDQKEEYOL 120  
 DB 61 ILNGSGFLTKGSPKLNDRADSRSLMDQGNPILITKLTIEDSDTYICEVEDQKEEYOL 120  
 QY 121 LVFGLTANSPTHLLOGSLTLTLESPPGSSPSVOCRSRPRGNIOGKTLVSQLELODSG 180  
 DB 121 LVFGLTANSPTHLLOGSLTLTLESPPGSSPSVOCRSRPRGNIOGKTLVSQLELODSG 180  
 QY 181 TWCTCTVLONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240  
 DB 181 TWCTCTVLONQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFTVEKLTGSGELMW 240

QY 241 QARRASSKSNITFDLKNKEVSVRVYTOPDKLQMGKKLPYLTLPOALPOYAGSGLTTLA 300  
 DB 241 QARRASSKSNITFDLKNKEVSVRVYTOPDKLQMGKKLPYLTLPOALPOYAGSGLTTLA 300  
 QY 301 LEAKTGLHGVNLVVMRATQLOKNLTCEVWGPTSPKLMSTLKENKAQVSKREKPVVY 360  
 DB 301 LEAKTGLHGVNLVVMRATQLOKNLTCEVWGPTSPKLMSTLKENKAQVSKREKPVVY 360  
 QY 361 LNPBAGMOCCLSDSGVLLSENIKVLPTWSTPV 394  
 DB 361 LNPBAGMOCCLSDSGVLLSENIKVLPTWSTPV 394

RESULT 86  
 AAR1042  
 ID AAR41042 standard; protein; 729 AA.  
 XX  
 AC AAR41042;  
 XX  
 DT 24-OCT-2003 (revised)  
 DT 25-MAR-2003 (revised)  
 DT 22-MAR-1994 (first entry)  
 XX  
 DE CD4-GBPH fusion protein.  
 XX  
 KM Merozoite; Glycophorin Binding Protein homologue; malaria; HIV; env;  
 KM human immunodeficiency virus; envelope glycoprotein; hybrid protein;  
 KM red blood cell; erythrocyte; AIDS.  
 XX  
 OS Homo sapiens.  
 OS Plasmodium falciparum.  
 OS Chimeric.  
 XX  
 FT Key Location/Qualifiers  
 FT Region 1..371  
 FT /note="residues 1-371 of CD4"  
 FT Region 372..729  
 FT /note="residues 70-427 of GBPH"  
 FT  
 PN WO9318160-A1.  
 PD 16-SEP-1993.  
 PF 10-MAR-1993; 93MO-GB000505.  
 XX  
 PR 11-MAR-1992; 92GB-00005276.  
 PR 08-JUL-1992; 92GB-00014481.  
 PR 24-JUL-1992; 92GB-00015829.  
 PR 16-SEP-1992; 92GB-00019562.  
 PR 03-MAR-1993; 93GB-00004311.  
 XX  
 PA (PREN/) PRENDERGAST K F.  
 PI Prendergast KF;  
 XX  
 DR WPI; 1993-303474/38.  
 PT Anti-viral fusion peptide(s) - comprise viral-binding component and  
 PT malaria merozoite red cell binding component, for treating e.g. HIV, and  
 PT hepatitis.  
 XX  
 PS Claim 7; Page 40-41; 69pp; English.  
 XX  
 CC The hybrid protein NH2-CD4(1-371)-GBPH(70-427)-COOH is a specifically  
 CC claimed example of a fusion protein of the invention; it comprises at  
 CC least part of the CD4 molecule fused to a peptide from a malarial  
 CC parasite merozoite protein with affinity for red blood cells. The fusion  
 CC protein can bind free HIV in the blood to red blood cells and  
 CC consequently reduce viral titre, prevent transmission of the virus and  
 CC improve safety of blood transfusions. (Updated on 25-MAR-2003 to correct  
 CC PN field.) (Updated on 24-OCT-2003 to standardise OS field)

Query Match 56.5%; Score 1929.5; DB 2; Length 729;  
 Best Local Similarity 63.9%; Pred. No. 4,5e-96;  
 Matches 434; Conservative 34; Mismatches 104; Indels 107; Gaps 18;

QY 24 OGNKVVLGKGGDTVELTCTASQKSIQFPMKNSNOIKIINGQSPFLTKGPSKLNDRADR 83  
 DB 1 OGNKVVLGKGGDTVELTCTASQKSIQFPMKNSNOIKIINGQSPFLTKGPSKLNDRADR 60  
 QY 84 RSLMDQGNFLLIKKLIKEDSDTYICEVDQKEVQLVFGILTANS DTHLQGSITLTL 143  
 DB 61 RSLMDQGNFLLIKKLIKEDSDTYICEVDQKEVQLVFGILTANS DTHLQGSITLTL 120  
 QY 144 ESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTCTVQLNQKVEFKRIDIVLVA 203  
 DB 121 ESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTCTVQLNQKVEFKRIDIVLVA 180  
 QY 204 FOKASIVYKKEGEOVEFSPLAFTVEKLTSGELMWOARRASSKSNITFDLKNKEVS 263  
 DB 181 FOKASIVYKKEGEOVEFSPLAFTVEKLTSGELMWOARRASSKSNITFDLKNKEVS 240  
 QY 264 KRVTDOPKLOMGKKLPYLTLPOALPOYAGSGLTTLAEAKTGKLOHGVNLVVMRATQLO 323  
 DB 241 KRVTDOPKLOMGKKLPYLTLPOALPOYAGSGLTTLAEAKTGKLOHGVNLVVMRATQLO 300  
 QY 324 KNLTCVWGPSPKMLSLKENKAQVSKREKPVVNLNPEAGMOCCLSDSGVLLSENI 383  
 DB 301 KNLTCVWGPSPKMLSLKENKAQVSKREKPVVNLNPEAGMOCCLSDSGVLLSENI 360  
 QY 384 IKVLPWSTPV-----PCPAPRPSCKDHTCPILGLGPSVFLPFPKPKDTL 430  
 DB 361 IKVLPWSTPVQYQADYSPRESRVLAEGSTSKNAKTL-----RKKQDTT 410  
 QY 431 MISRPEVTCVVVDVSHEDPEVKFMNYYDGEVHNAKTKPREQYNST----- 478  
 DB 411 LRSADPEQ--IMKMAADPEYRKHLNVLQILNN--TDPNDELTSADPEQIMKAYAA 466  
 QY 479 ---YRVSVLYLHQMNLNGKEYKCVSNKALPAPLEKTSISKAKQPREPQYTLTPPSND 535  
 DB 467 DPEYR--KHLNVLQY--ILNNTDNDVESSADP--EQQIMKA-----YAADP-- 508  
 QY 536 ELTKQVSLTCLVKGFPSPDIAVEMESNQOPEN-----NYKTPPVL-----DS 580  
 DB 509 EYRKHNVLQILNNTDND---ELETSADPEQIMKAYAADPEYRKHNVLQILNHTD 565  
 QY 581 GSEFLYSKLTVDKSRMQGNVFS-----SVMEALNHYTKSLSPQLQD 629  
 DB 566 SS-----EVEITSADPEQIMKAYAADPEYRKHNVLQIL--NHTDSSEV----- 608  
 QY 630 ETCAENQDELQGLWTTDP 648  
 DB 609 ETSAD-PEQIMKAYAADP 626

RESULT 87  
 AAR15149  
 ID AAR15149 standard; protein; 458 AA.  
 XX  
 AC AAR15149;  
 XX  
 DT 25-MAR-2003 (revised)  
 DT 24-FEB-1992 (first entry)  
 XX  
 DE CD4 coordinate system.  
 XX  
 KM Gelolin; fusion protein; diagnosis; AIDS.  
 XX  
 OS Homo sapiens.  
 XX  
 FT Key Location/Qualifiers  
 FT Domain 1..25  
 FT /label= hydrophobic/secretory\_signal  
 FT Domain 26..132

```

FT FT /label= first_Ig-related_domain
FT FT /note= "extracellular"
FT FT Diulfide-bond
FT FT 41. .109
FT FT 133. .202
FT FT Domain
FT FT /label= second_Ig-related_domain
FT FT /note= "extracellular"
FT FT Diulfide-bond
FT FT 155. .184
FT FT 203. .318
FT FT Domain
FT FT /label= third_Ig-related_domain
FT FT /note= "extracellular"
FT FT 319. .395
FT FT /label= fourth_Ig-related_domain
FT FT /note= "extracellular"
FT FT Diulfide-bond
FT FT 328. .370
FT FT 396. .416
FT FT Domain
FT FT /label= hydrophobic/membrane_sequence
FT FT 417. .458
FT FT /label= very_hydrophilic/intracytoplasmic
FT FT Domain
FT FT WO9117170-A.
FT FT 14-NOV-1991.
FT FT 90US-00520368.
FT FT 90US-00520368.
FT FT 04-MAY-1990; 90US-00520368.
FT FT 04-MAY-1990; 90US-00520368.
FT FT (BIOJ ) BIOGEN INC.
FT FT Pepinsky RB, Rosa MD, Stossel TP;
FT FT MPI; 1991-353711/48.
FT FT N-PSDB; AAQ14931.
FT FT New multi-meric and hetero-multi-meric gelsolin fusion constructs - used
FT FT to treat and diagnose AIDS, ARC and HIV infection.
FT FT Disclosure; fig 3A-3D; 131pp; English.
FT FT The CD4 polypeptides useful in the constructs include all CD4
FT FT polypeptides which bind to or otherwise inhibit gp120 and gp160. These
FT FT include fragments lacking the transmembrane domain. In particular it is
FT FT CD4 1-111; 1-Cys111; 1-Cys180; 1-181, 1-183; 1-187; 1-345 or 1-375 (from
FT FT mature protein). See also AAQ14931-35 and AAR15151. (Updated on 25-MAR-
FT FT 2003 to correct PA field.)
FT FT
FT FT Sequence 458 AA:
SQ
Query Match 56.3%; Score 1921; DB 2; Length 458;
Best Local Similarity 95.2%; Pred. No. 8e-96; 12; Indels 0; Gaps 0;
Matches 375; Conservative 7; Mismatches 12;
QY 1 MNRGVPFRHLLVQLALLPAATQGNKTVLGKGGDTVELTCTASQKSIQFHWKNSNQIK 60
DB 1 MNRGVPFRHLLVQLALLPAATQGNKTVLGKGGDTVELTCTASQKSIQFHWKNSNQIK 60
QY 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIIKVKLIEDSPYICEVEDQKEEYOL 120
DB 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNFPLIIKVKLIEDSPYICEVEDQKEEYOL 120
QY 121 LVFGITANSPTHLLOGSLTLTLESPGSPSVOCRSRGNKIOGKTLVSQLELDQSG 180
DB 121 LVFGITANSPTHLLOGSLTLTLESPGSPSVOCRSRGNKIOGKTLVSQLELDQSG 180
QY 121 LVFGITANSPTHLLOGSLTLTLESPGSPSVOCRSRGNKIOGKTLVSQLELDQSG 180
DB 121 LVFGITANSPTHLLOGSLTLTLESPGSPSVOCRSRGNKIOGKTLVSQLELDQSG 180
QY 181 TWCTVNLONOKKVEFKIDIVLAFQKASSIVYKKEGEVFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTVNLONOKKVEFKIDIVLAFQKASSIVYKKEGEVFSPLAFTVEKLTGSGELMW 240
QY 181 TWCTVNLONOKKVEFKIDIVLAFQKASSIVYKKEGEVFSPLAFTVEKLTGSGELMW 240
DB 181 TWCTVNLONOKKVEFKIDIVLAFQKASSIVYKKEGEVFSPLAFTVEKLTGSGELMW 240
QY 241 QAERASSSSKSWITPPLKKEVSVKRVTPDKLQNGKKLPLHLTLTPOALPOYAGSGNTLA 300
DB 241 QAERASSSSKSWITPPLKKEVSVKRVTPDKLQNGKKLPLHLTLTPOALPOYAGSGNTLA 300
QY 301 LEAKTGKHQEVNLVVMRATOLQKNLTCEVWGPTSPKLMLSLKENKAKVSKKKPVWV 360

```

```

DB 301 LEAKTGKHQEVNLVVMRATOLQKNLTCEVWGPTSPKLMLSLKENKAKVSKKKPVWV 360
QY 361 LNPEAGWQCILSDSGQVLESNIKVLPTWSTPV 394
DB 361 LNPEAGWQCILSDSGQVLESNIKVLPTWSTPV 394
RESULT 88
AAR41041
ID AAR41041 standard; protein; 942 AA.
AC AAR41041;
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 22-MAR-1994 (first entry)
DE CD4-GBP130 fusion protein.
XX KM Merozoite; Glycophorin Binding Protein 130; malaria; HIV; env;
XX KM human immunodeficiency virus; envelope glycoprotein; hybrid protein;
XX KM red blood cell; erythrocyte; AIDS.
XX OS Homo sapiens.
XX OS Plasmodium falciparum.
XX OS Chimeric.
XX FH Key
XX FH Region
XX FH Location/Qualifiers
FT 1. .371
FT /note= "residues 1-371 of CD4"
FT Region
FT /note= "residues 201-774 of GBP130"
XX PN WO9318160-A1.
XX PD 16-SEP-1993.
XX PF 10-MAR-1993; 93WO-GB000505.
XX PR 11-MAR-1992; 92GB-00005276.
XX PR 08-JUL-1992; 92GB-00014481.
XX PR 24-JUL-1992; 92GB-00015829.
XX PR 16-SEP-1992; 92GB-00019562.
XX PR 03-MAR-1993; 93GB-00004311.
XX PA (PREN/) PRENDERGAST K F.
XX PI Prendergast KF;
XX WP; 1993-303474/38.
XX PT Anti-viral fusion peptide(s) - comprise viral-binding component and
XX PT malaria merozoite red cell binding component, for treating e.g. HIV, and
XX PT hepatitis.
XX PS Claim 2; Page 35-37; 69pp; English.
XX CC The hybrid protein NH2-CD4(1-371)-GBP130(201-774)-COOH is a specifically
XX CC claimed example of a fusion protein of the invention: it comprises at
XX CC least part of the CD4 molecule fused to a peptide from a malarial
XX CC parasite merozoite protein with affinity for red blood cells. The fusion
XX CC protein can bind free HIV in the blood to red blood cells and
XX CC consequently reduce viral titre, prevent transmission of the virus and
XX CC improve safety of blood transfusions. (Updated on 25-MAR-2003 to correct
XX CC PN field.) (Updated on 24-OCT-2003 to standardise OS field)
SQ
Sequence 942 AA:
Query Match 55.9%; Score 1908.5; DB 2; Length 942;
Best Local Similarity 84.5%; Pred. No. 8e-95;
Matches 386; Conservative 14; Mismatches 32; Indels 25; Gaps 5;

```

```

QY 24 QGNKVVLTGKKGDTVELTCTASQKKSIOFHMKNSNOIKILGNQSFLLTKGPSKLNDRADSR 83
DB 1 QGNKVVLTGKKGDTVELTCTASQKKSIOFHMKNSNOIKILGNQSFLLTKGPSKLNDRADSR 60
QY 84 RSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEVLVFGLTANSPTHLLOQSITLTL 143
DB 61 RSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEVLVFGLTANSPTHLLOQSITLTL 120
QY 144 ESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOQKKEFKIDIVLVA 203
DB 121 ESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOQKKEFKIDIVLVA 180
QY 204 FQKASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOABRASSSKSWITFDLKNKEVS 263
DB 181 FQKASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOABRASSSKSWITFDLKNKEVS 240
QY 264 KRVTDPKLOMGKKLPLHLTLPQALPQYAGSGNLTALBAKTGKLHGVNLVVMRATOLQ 323
DB 241 KRVTDPKLOMGKKLPLHLTLPQALPQYAGSGNLTALBAKTGKLHGVNLVVMRATOLQ 300
QY 324 KNITCEVWGPTSPKMLSLKENKEAKVSKREKPVVTLNPEAGMOCCLSDSGQVLLSEN 383
DB 301 KNITCEVWGPTSPKMLSLKENKEAKVSKREKPVVTLNPEAGMOCCLSDSGQVLLSEN 360
QY 384 IKVLPWTSTVPCPAPPEPKSCDHTHTCEPBLGGPSVFLFPKPKDPTLMISRTPEVTCVVY 443
DB 361 IKVLPWTSTPV---SQKPESTSTRSN-----EVKIRAAASNOETL-TSADPEGO--IM 406
QY 444 DVSHEDPEVKENMYV-----DGVENHNAKTK 469
DB 407 REYAADPEYRKHLIEFYKILNTDPNDEVERRADNK 443

RESULT 89
AAP90833
ID AAP90833 standard; protein; 384 AA.
XX
AC AAP90833;
XX
DT 25-MAR-2003 (revised)
DT 01-AUG-1990 (first entry)
XX
DE Amino acid sequence of a soluble T4-like (ST4) polypeptide encoded by a
DE portion of clone p199-7 (PL mutet.rst4).
XX
KW HIV; immunotherapeutic; prophylactic; soluble T4-like polypeptide;
KW diagnostic; p199-7 (PL mutet.rst4).
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Misc-difference 10 /note= "initiating Met"
XX
XX MO8901940-A.
XX
XX PD 09-MAR-1989.
XX
XX PF 01-SEP-1988; 88WO-US002940.
XX
XX PR 04-SEP-1987; 87US-00094322.
XX
XX PR 07-JAN-1988; 88US-00141649.
XX
XX PA (BIOI ) BIOGEN INC.
XX
XX PI Fisher RA, Gilbert W, Sato VL, Flavell RA, Maraganore JM;
XX
XX DR WPI: 1989-085519/11.
XX
XX DR N-PSDB; AAN90643.
XX
XX PT DNA sequences coding for soluble T4-like polypeptide(s) - used in
XX immunotherapeutic and immunosuppressive compans. and for preventing,
XX treating or detecting AIDS.
XX

```

```

XX
PS Disclosure; Page 7; 207pp; English.
XX
CC It is the protein sequence encoded by the rst4 sequence. It is claimed in
CC the patent. It is useful in immunotherapeutic, prophylactic and
CC diagnostic compans. It can be used to purify HIV from a sample. (Updated
CC on 25-MAR-2003 to correct PR field.)
XX
SQ Sequence 384 AA;
Query Match 55.9%; Score 1908; DB 1; Length 384;
Best Local Similarity 99.7%; Pred. No. 3,4e-95;
Matches 370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 24 QGNKVVLTGKKGDTVELTCTASQKKSIOFHMKNSNOIKILGNQSFLLTKGPSKLNDRADSR 83
DB 11 QGNKVVLTGKKGDTVELTCTASQKKSIOFHMKNSNOIKILGNQSFLLTKGPSKLNDRADSR 70
QY 84 RSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEVLVFGLTANSPTHLLOQSITLTL 143
DB 71 RSLMDQGNFPLIINKLIKEDSDTYICEVEDQKEEVLVFGLTANSPTHLLOQSITLTL 130
QY 144 ESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOQKKEFKIDIVLVA 203
DB 131 ESPPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOQKKEFKIDIVLVA 190
QY 204 FQKASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOABRASSSKSWITFDLKNKEVS 263
DB 191 FQKASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWOABRASSSKSWITFDLKNKEVS 250
QY 264 KRVTDPKLOMGKKLPLHLTLPQALPQYAGSGNLTALBAKTGKLHGVNLVVMRATOLQ 323
DB 251 KRVTDPKLOMGKKLPLHLTLPQALPQYAGSGNLTALBAKTGKLHGVNLVVMRATOLQ 310
QY 324 KNITCEVWGPTSPKMLSLKENKEAKVSKREKPVVTLNPEAGMOCCLSDSGQVLLSEN 383
DB 311 KNITCEVWGPTSPKMLSLKENKEAKVSKREKPVVTLNPEAGMOCCLSDSGQVLLSEN 370
QY 384 IKVLPWTSTPV 394
DB 371 IKVLPWTSTPV 381

RESULT 90
AAR41043
ID AAR41043 standard; protein; 1786 AA.
XX
AC AAR41043;
XX
DT 24-OCT-2003 (revised)
DT 25-MAR-2003 (revised)
DT 22-MAR-1994 (first entry)
XX
DE CD4-EBA175 fusion protein.
XX
XX Merzoite; Erythrocyte Binding Antigen 175; malaria; HIV; env;
XX human immunodeficiency virus; envelope glycoprotein; hybrid protein;
XX red blood cell; erythrocyte; AIDS; molecular machine.
XX
XX OS Homo sapiens.
XX
XX OS Plasmodium falciparum.
XX
XX OS Chimeric.
XX
XX FH Key Location/Qualifiers
XX
XX FT Region 1..371
XX
XX FT Region /note= "residues 1-371 of CD4"
XX
XX FT Region 372..1786
XX
XX FT /note= "residues 20-1435 of EBA-175"
XX
XX MO9318160-A1.
XX
XX PN 16-SEP-1993.
XX

```

```

PF 10-MAR-1993; 93WO-GB000505.
XX
XX 11-MAR-1992; 92GB-00005276.
PR 08-JUL-1992; 92GB-00014481.
PR 24-JUL-1992; 92GB-00015829.
PR 16-SEP-1992; 92GB-00019562.
PR 03-MAR-1993; 93GB-00004311.
XX
PA (PREN/) PRENDERGAST K F.
XX
PI Prendergast KF;
XX
DR WPI; 1993-303474/38.
XX
PT Anti-viral fusion peptide(s) - comprise viral-binding component and
PT malaria merozoite red cell binding component, for treating e.g. HIV, and
PT hepatitis.
XX
PS Claim 9; Page 44-47; 69pp; English.
XX
CC The hybrid protein NH2-CD4(1-371)-EBA175(20-1435)-COOH is a specifically
CC claimed example of a fusion protein of the invention; it comprises at
CC least part of the CD4 molecule fused to a peptide from a malaria
CC parasite merozoite protein with affinity for red blood cells. The fusion
CC protein can bind free HIV in the blood to red blood cells and
CC consequently reduce viral titre, prevent transmission of the virus and
CC improve safety of blood transfusions. (updated on 25-MAR-2003 to correct
CC PN field.) (updated on 24-OCT-2003 to standardise OS field)
XX
SQ Sequence 1786 AA;
XX
Query Match 55.9%; Score 1908; DB 2; Length 1786;
Best Local Similarity 99.7%; Pred. No. 1.7e-94;
Matches 370; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 24 QGNKVLGKGGDTVELTCTASQKSIQPHMKNQIKILGNQGSFLTKGPSKLNDRADSR 83
DB 1 QGNKVLGKGGDTVELTCTASQKSIQPHMKNQIKILGNQGSFLTKGPSKLNDRADSR 60
QY 84 RSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSSTHLLQGSFLTL 143
DB 61 RSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSSTHLLQGSFLTL 120
QY 144 ESPGSSPSVQCRSPRKNIOGGKTLVSQLELDSDGTCTVYLQNKQVEFKIDIVVLA 203
DB 121 ESPGSSPSVQCRSPRKNIOGGKTLVSQLELDSDGTCTVYLQNKQVEFKIDIVVLA 180
QY 204 FOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMQAERASSSKSWITFDLKNKEYSV 263
DB 181 FOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMQAERASSSKSWITFDLKNKEYSV 240
QY 264 KRYTQDPKLGKGLPLHLTLPLQALPOYAGSGNLTALAEAKTGKHQEVNLVVMRAATOLQ 323
DB 241 KRYTQDPKLGKGLPLHLTLPLQALPOYAGSGNLTALAEAKTGKHQEVNLVVMRAATOLQ 300
QY 324 KNLTCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVNLNPEAGMOCILSDSGVLLSEN 383
DB 301 KNLTCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVNLNPEAGMOCILSDSGVLLSEN 360
QY 384 IKVLPWTWSTPV 394
DB 361 IKVLPWTWSTPV 371
XX
RESULT 91
ID AAP96151 standard; protein; 434 AA.
AC AAP96151;
XX
XX 25-MAR-2003 (revised)
DT 24-DEC-1990 (first entry)
XX

```

```

DE Sequence of a fusion of the herpes virus and N-terminal 27 residues to
DE the putative mature N-terminus of CD4T.
XX
XX gp120; HIV; therapy; adhesion; antiviral.
XX
XX Synthetic.
XX
XX EP14317-A.
XX
XX 03-MAY-1989.
XX
XX 03-OCT-1988; 88EP-00309194.
XX
XX 02-OCT-1987; 87US-00104329.
XX
XX 28-SEP-1988; 88US-00250785.
XX
XX (GETH ) GENENTECH INC.
XX
XX Capon DJ, Gregory TJ;
XX
XX WPI; 1989-131855/18.
XX
XX N-PSDB; AAN90778.
XX
XX Compans. contg. adhesion variants - useful in therapy and diagnostics,
XX e.g. CD4 variants which are therapeutically useful for treating human
XX immuno-deficiency virus.
XX
XX Example; Fig 2a-2c; 36pp; English.
XX
XX A nucleic acid encoding an AA sequence variant of an adhesion is claimed.
XX Adhesion variants may be a fusion of a CD4 polypeptide and a polypeptide
XX different from CD4. Used therapeutically or as diagnostic reagents for
XX the assay of adhesions or their ligands, or for purificn. of the adhesions.
XX The CD4 adhesion variants are used for antiviral or immunomodulatory
XX therapy, in partic. treatment of HIV infection. (updated on 25-MAR-2003
XX to correct PI field.)
XX
SQ Sequence 434 AA;
XX
Query Match 55.8%; Score 1904; DB 1; Length 434;
Best Local Similarity 99.7%; Pred. No. 6.3e-95;
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 24 QGNKVLGKGGDTVELTCTASQKSIQPHMKNQIKILGNQGSFLTKGPSKLNDRADSR 83
DB 56 QGNKVLGKGGDTVELTCTASQKSIQPHMKNQIKILGNQGSFLTKGPSKLNDRADSR 115
QY 84 RSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSSTHLLQGSFLTL 143
DB 116 RSLMDQGNFPLIINKLKIEDSDTYICEVEDQKEVQLVFGLTANSSTHLLQGSFLTL 175
QY 144 ESPGSSPSVQCRSPRKNIOGGKTLVSQLELDSDGTCTVYLQNKQVEFKIDIVVLA 203
DB 176 ESPGSSPSVQCRSPRKNIOGGKTLVSQLELDSDGTCTVYLQNKQVEFKIDIVVLA 235
QY 204 FOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMQAERASSSKSWITFDLKNKEYSV 263
DB 236 FOKASSIYKKEGEQVEFSPLAFTVEKLTGSGELMQAERASSSKSWITFDLKNKEYSV 295
QY 264 KRYTQDPKLGKGLPLHLTLPLQALPOYAGSGNLTALAEAKTGKHQEVNLVVMRAATOLQ 323
DB 296 KRYTQDPKLGKGLPLHLTLPLQALPOYAGSGNLTALAEAKTGKHQEVNLVVMRAATOLQ 355
QY 324 KNLTCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVNLNPEAGMOCILSDSGVLLSEN 383
DB 356 KNLTCEVWGPTSPKLMSTLKENKEAKVSKREKPVWVNLNPEAGMOCILSDSGVLLSEN 415
QY 384 IKVLPWTWSTP 393
DB 416 IKVLPWTWSTP 425
XX
RESULT 92

```

AA093557  
ID AA093557 standard; protein; 434 AA.  
AC  
XX AAP93557;  
XX  
XX 25-MAR-2003 (revised)  
DT 03-OCT-2002 (revised)  
DT 06-JUN-1990 (first entry)  
XX  
XX Fusion of the herpes gD leader and N-terminal 27 residues to mature N-terminus of CD4T.  
DE  
XX Adhesion variant; CD4T; soluble CD4; truncated CD4; antiviral; HIV;  
KM herpes gD; immunomodulatory; diagnostic.  
XX  
XX Homo sapiens.  
OS  
XX MO8902922-A.  
PN  
XX 06-APR-1989.  
PD  
XX 03-OCT-1988; 88WO-US003414.  
PF  
XX 02-OCT-1987; 87US-00104329.  
PR 28-SEP-1988; 88US-00250785.  
PR  
XX (GETH ) GENENTECH INC.  
PA  
XX Capon DJ, Gregory TJ;  
PI  
XX WPI; 1989-114397/15.  
DR N-PSDB; AAN90735.  
XX  
XX New nucleic acid sequences encoding adhesion, esp. CD 4, variants -  
PT partic. with trans-membrane domain inactivated or fused to other peptide,  
PT useful esp. for treating HIV infections.  
XX  
XX Fig 2A-2C; pp. 5/13-8/13; 78pp; English.  
PS  
XX CD4T is a truncated or soluble variant of CD4. CD4T fusion proteins can  
CC have antiviral and immunomodulatory activity and are esp. useful for  
CC treating HIV infections, regardless of genetic variation within the  
CC virus. CD4T fusion proteins, and antibodies raised against them, can also  
CC be used diagnostically for assaying adhesions and their ligands. (Updated  
CC on 03-OCT-2002 to add missing OS field.) (Updated on 25-MAR-2003 to  
CC correct PR field.) (Updated on 25-MAR-2003 to correct PA field.)  
CC  
XX  
XX  
SQ Sequence 434 AA;  
Query Match 55.8%; Score 1904; DB 1; Length 434;  
Best Local Similarity 99.7%; Pred. No. 6.3e-95;  
Matches 369; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 324 KNLTCVWGPTSPKMLSLKLENKEAKVSKREKPVVILNPBAQMOCLLSDSGQVLLSESN 383  
DB |||||  
DB 356 KNLTCVWGPTSPKMLSLKLENKEAKVSKREKAVVILNPBAQMOCLLSDSGQVLLSESN 415  
QY 384 IKVLPWSTP 393  
DB |||||  
DB 416 IKVLPWSTP 425  
RESULT 93  
AA07721  
ID AA07721 standard; protein; 375 AA.  
XX  
XX AA07721;  
AC  
XX 25-MAR-2003 (revised)  
DT 18-FEB-1991 (first entry)  
XX  
XX Recombinant soluble (rs) T4.  
DE  
XX  
XX Metal-binding site.  
KM  
XX  
XX Homo sapiens.  
OS  
XX MO9012803-A.  
PN  
XX 01-NOV-1990.  
PD  
XX 14-APR-1989; 89US-00338991.  
PF  
XX 14-APR-1989; 89US-00338991.  
PR  
XX 14-APR-1989; 89US-00338991.  
PR  
XX (BIOI ) BIOGEN INC.  
PA (BIOI ) BIOGEN INC.  
PA  
XX Staples MA, Pargellis CA;  
PI  
XX  
XX WPI; 1990-348421/46.  
DR  
XX  
XX Purifying protein having surface metal-binding amino acid residues -  
PT using an immobilised metal affinity chromatography resin.  
PT  
XX  
XX Disclosure; Fig 2; 36pp; E.  
PS  
XX  
XX The rs T4 can be purified from a crude sample (updated on 25-MAR-2003 to  
CC correct PA field.)  
CC  
XX  
XX  
SQ Sequence 375 AA;  
Query Match 55.5%; Score 1894; DB 2; Length 375;  
Best Local Similarity 99.5%; Pred. No. 1.9e-94;  
Matches 367; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 326 LTCEVWGPTSPKMLSLKLENKAKVSKREKRPVWVNLPEAGMOCLLSDSGVLLSENIK 385  
 Db 301 LTCEVWGPTSPKMLSLKLENKAKVSKREKRPVWVNLPEAGMOCLLSDSGVLLSENIK 360  
 Qy 386 VLPTWSTPV 394  
 Db 361 VLPTWSTPV 369

## RESULT 94

AAW41376  
 ID AAW41376 standard; peptide; 433 AA.

AC AAW41376;

DT 28-MAY-1998 (first entry)

DE Human CD4.

KW Antibody; CD4; passive immunity; HIV type 1; HIV type 2; HIV infection;  
 KX sultan immunodeficiency virus; SIV; AIDS; therapy; HIV gp120.

OS Homo sapiens.

PN MO9746697-A2.

PD 11-DEC-1997.

PF 03-JUN-1997; 97WO-US009449.

PR 03-JUN-1996; 96US-00657149.

PR 28-FEB-1997; 97US-00808374.

PR 02-JUN-1997; 97US-00867149.

PA (UNBI-) UNITED BIOMEDICAL INC.

PI Wang CY;

DR WPI: 1998-042204/04.

PS Disclosure: Page 122-123; 140pp; English.

This sequence represents human CD4, and was used to isolate the antibody (Ab) of the invention. The Ab, preferably a M2 or B13 Ab, has the following characteristics: (a) binds rscD4; (b) binds CD4 expressing cells in an immunofluorescence assay, where the binding pattern is in the shape of "caps", when examined with a high resolution fluorescence microscope; (c) blocks the binding of HIV glycoprotein 120 (gp120) to CD4 expressing cells; (d) binds CD4 expressing cells previously bound with gp120; and (e) neutralises HIV primary isolates in an in vitro microplaque assay at a concentration of less than 10 mu g/ml. Preferably at a concentration in the range of 0.01-10 mu g/ml for 50% neutralisation and 0.1-35 mu g/ml for 90% neutralisation. The Ab can be used to provide passive immunity to HIV in a mammal, when administered parenterally, specifically all clades of HIV type 1, and from diverse primary isolates of HIV type 2 and sultan immunodeficiency virus (SIV). The Ab is prophylactic and therapeutic for HIV infection and all stages of AIDS because it prevents replicative infection of host cells both before and after HIV gp120 has bound to the host cell antigen complex comprising CD4 on the surface of CD4 positive lymphocytes, thus it is capable of preventing HIV infection and retarding the spread of the virus to uninfected cells. It is also uniquely useful because it inhibits infection following binding of HIV to CD4 expressing cells

Sequence 433 AA;

Query Match 55.4%; Score 1892; DB 2; Length 433;  
 Best Local Similarity 99.5%; Pred. No. 2.8e-94;  
 Matches 367; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 26 NKVVLGKKGDVVELTCTASQKSIQFHMKNNSNOIKILGNQSPFLTKGPKLNDRAISR 85  
 Db 1 NKVVLGKKGDVVELTCTASQKSIQFHMKNNSNOIKILGNQSPFLTKGPKLNDRAISR 60  
 Qy 86 LMDQGNFPLIIKNLIKIDSDTYICEVEDQKEVOLLVFGLTANSDFHLLQGSITLTL 145  
 Db 61 LMDQGNFPLIIKNLIKIDSDTYICEVEDQKEVOLLVFGLTANSDFHLLQGSITLTL 120  
 Qy 146 PGSSPSVQCRSPRGKNIQGGKTLISVQLQDSGTCTVLQNKVEFFIDIVLAFQ 205  
 Db 121 PGSSPSVQCRSPRGKNIQGGKTLISVQLQDSGTCTVLQNKVEFFIDIVLAFQ 180  
 Qy 206 KASSIVYKKEGEQVEFFPLAFVEKLTGSGELMWQERASSKSWITFDLKNVEVSR 265  
 Db 181 KASSIVYKKEGEQVEFFPLAFVEKLTGSGELMWQERASSKSWITFDLKNVEVSR 240  
 Qy 266 VTQDPKIQMGKQLPLHLTLFQALPQVAGSGLTLALEAKTGKHQEVNLVVMRATOLQKN 325  
 Db 241 VTQDPKIQMGKQLPLHLTLFQALPQVAGSGLTLALEAKTGKHQEVNLVVMRATOLQKN 300  
 Qy 326 LTCEVWGPTSPKMLSLKLENKAKVSKREKRPVWVNLPEAGMOCLLSDSGVLLSENIK 385  
 Db 301 LTCEVWGPTSPKMLSLKLENKAKVSKREKRPVWVNLPEAGMOCLLSDSGVLLSENIK 360  
 Qy 386 VLPTWSTPV 394  
 Db 361 VLPTWSTPV 369

## RESULT 95

AAV39824  
 ID AAV39824 standard; protein; 369 AA.

AC AAV39824;

DT 03-DEC-1999 (first entry)

DE Soluble human T4 protein.

KW Soluble T4 protein; sT4; human; HIV; binding inhibitor; T4+ cell; AIDS;  
 KX vaccine; immunisation; therapy.

OS Homo sapiens.

PN US5958678-A.

PD 28-SEP-1999.

PF 12-DEC-1994; 94US-00354452.

PR 21-AUG-1986; 86US-00898587.

PR 11-JUN-1991; 91US-00713564.

PR 06-JUL-1992; 92US-00909021.

PA (UYCO) UNIV COLUMBIA NEW YORK.

PI McDougal JS, Weiss R, Axel R, Littman DR, Maddon PJ, Chess L;

DR WPI: 1999-561025/47.

PS Claim 1; Col 51-53; 58pp; English.

This sequence represents the soluble human T4 protein of the invention. The soluble human T4 protein blocks the binding of HIV to T4+ cells and is therefore useful for the treatment of AIDS. Monoclonal antibodies against the T4 protein may be used as vaccines for immunising subjects against AIDS

Sequence 369 AA;

Query Match 55.4%; Score 1891; DB 2; Length 369;  
 Best Local Similarity 99.7%; Pred. No. 2.7e-94;  
 Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 27 KVLGKKGDVLTCTASQKKSIOFHWKNSNOIKILNGSFLTKGPSKLNDRADSRSL 86  
 CC KVLGKKGDVLTCTASQKKSIOFHWKNSNOIKILNGSFLTKGPSKLNDRADSRSL 61  
 DB 2 KVLGKKGDVLTCTASQKKSIOFHWKNSNOIKILNGSFLTKGPSKLNDRADSRSL 61

QY 87 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFGLTANSDFHLLQGSLTLLSESP 146  
 DB WDOGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFGLTANSDFHLLQGSLTLLSESP 121

QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTVLONOKKVEFKIDIVLAFOK 206  
 DB PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTVLONOKKVEFKIDIVLAFOK 181

QY 207 ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSSKSWITFDLKNKEVSKRV 266  
 DB ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSSKSWITFDLKNKEVSKRV 241

QY 267 TODPKLQMGKKLPLHLTLPQALPOYAGSGNLTALAEAKTGKJHOEVNLVVMRATQLOK 326  
 DB 242 TOPDKLQMGKKLPLHLTLPQALPOYAGSGNLTALAEAKTGKJHOEVNLVVMRATQLOK 301

QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCCLSDSGOVLLESNIRY 386  
 DB 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCCLSDSGOVLLESNIRY 361

QY 387 LPTWSTPV 394  
 DB 362 LPTWSTPV 369

RESULT 96  
 ID AAY88327 standard; protein; 369 AA.  
 AC AAY88327;  
 XX 14-JUL-2000 (first entry)  
 DT 14-JUL-2000 (first entry)  
 DE T4 glycoprotein amino acid sequence.  
 XX  
 KW sT4; glycoprotein; human immunodeficiency virus; HIV; block binding;  
 KM AIDS; treatment; inhibic; cell to cell spread; infection; fusion.  
 XX  
 OS Mammalia.  
 XX  
 PN USS126433-A.  
 XX 30-JUN-1992.  
 PD 30-JUN-1992.  
 PF 23-OCT-1987; 87US-00114244.  
 XX 21-AUG-1986; 86US-00898587.  
 PR  
 XX (UYCO ) UNIT COLUMBIA NEW YORK.  
 PA  
 PI Maddon PJ, Cheese L, Axel R, Weiss R, Littman DR, McDougal JS;  
 XX WPI; 2000-348913/30.  
 DR  
 XX Soluble T4 glycoprotein useful for prevention and treatment of acquired  
 PT immunodeficiency syndrome and for screening inhibitors of human  
 PT immunodeficiency viral binding.  
 XX  
 PS Claim 1; Col 54; 64pp; English.  
 CC This sequence represents the amino acid sequence of glycosylated sT4  
 CC glycoprotein. Human immunodeficiency virus (HIV) uses sT4 as a target  
 CC receptor on T cells. The invention relates to glycosylated sT4 which  
 CC functions by blocking the binding of HIV to T4 target cells, and can be  
 CC used for the prophylaxis and treatment of AIDS patients. Administration

CC of sT4 effectively inhibits the cell to cell spreading of HIV infection  
 CC and also the fusion of HIV-infected T4 cells and non-infected T4 cells.  
 CC The administration of T4 alleviates several symptoms associated with  
 CC AIDS, and prevents the occurrence of new pathological changes. The sT4  
 CC glycoprotein is useful for the prophylaxis and treatment of patients with  
 CC AIDS. It is also useful as a reagent to identify natural, synthetic or  
 CC recombinant molecules which act as therapeutic agents or inhibitors of  
 CC T4+ cell interactions and in diagnostic assays for detection T4 proteins  
 CC or molecules

QY 27 KVLGKKGDVLTCTASQKKSIOFHWKNSNOIKILNGSFLTKGPSKLNDRADSRSL 86  
 DB 2 KVLGKKGDVLTCTASQKKSIOFHWKNSNOIKILNGSFLTKGPSKLNDRADSRSL 61

QY 87 WDOGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFGLTANSDFHLLQGSLTLLSESP 146  
 DB WDOGNFPLIKNLKIEDSDTYICEVEDQKEVOLVFGLTANSDFHLLQGSLTLLSESP 121

QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTVLONOKKVEFKIDIVLAFOK 206  
 DB PGSSPSVQCRSPRGKNIQGGKTLVSQLELDGSGTWCTVLONOKKVEFKIDIVLAFOK 181

QY 207 ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSSKSWITFDLKNKEVSKRV 266  
 DB ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSSKSWITFDLKNKEVSKRV 241

QY 267 TODPKLQMGKKLPLHLTLPQALPOYAGSGNLTALAEAKTGKJHOEVNLVVMRATQLOK 326  
 DB 242 TOPDKLQMGKKLPLHLTLPQALPOYAGSGNLTALAEAKTGKJHOEVNLVVMRATQLOK 301

QY 327 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCCLSDSGOVLLESNIRY 386  
 DB 302 TCEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPNPEAGMOCCLSDSGOVLLESNIRY 361

QY 387 LPTWSTPV 394  
 DB 362 LPTWSTPV 369

RESULT 97  
 ID AAP93528 standard; protein; 370 AA.  
 AC AAP93528;  
 XX 25-MAR-2003 (revised)  
 DT 04-JUN-1990 (first entry)  
 DE Human soluble CD4 protein (T4ex1) encoded by T4 SECI cDNA.  
 XX  
 KW Human soluble CD4 protein; T4 SECI cDNA; T4ex1;  
 KM HIV gp120 envelope protein; T-lymphocyte.  
 XX  
 OS Homo sapiens.  
 XX  
 PN Key Location/Qualifiers  
 FH Key  
 FT Misc-difference 129  
 FT Misc-difference 163  
 FT Misc-difference 163  
 FT Misc-difference 243  
 FT Misc-difference 370  
 XX  
 PN W08903222-A.  
 XX 20-APR-1989.  
 PD 20-APR-1989.  
 PF 05-OCT-1988; 88WO-US003454.



XX 08-OCT-1987; 87US-00106185.  
PR 14-JAN-1988; 88US-00144313.  
PR 14-JUN-1988; 88US-00206585.  
PR 11-JUL-1988; 88US-00217475.  
XX  
PA (DANA-) DANA-FABER CANCER.  
XX  
PI Reinherz E, Huseby R, Droski J, Richardson N;  
XX WPI; 1989-122898/17.  
DR N-PSDB; AAN90764.  
XX  
PT Soluble human CD4 fragments - capable of binding HIV gp 120 envelope  
PT protein, used for diagnosis, treatment or prophylaxis of HIV infection.  
XX  
PS Fig 1; pages 1/11-3/11; 106pp; English.  
XX  
CC Misc-difference in the features table represent where a modification in  
CC cDNA and in the encoded CD4 protein has been made. A soluble human CD4  
CC fragment capable of binding to the gp. 120 envelope protein of HIV is  
CC claimed. Such a fragment should include none of the hydrophobic trans-  
CC membrane region of naturally-occurring CD4, or contain a portion of the  
CC hydrophobic region which is sufficiently short it does not prevent  
CC solubilization. Also claimed are modified soluble CD4 fragments which  
CC differ by subseqn. or addn. to this sequence (AAP93528) by at least one  
CC AA. Such fragments pref. consist of: the first 182, 369, 549, 729 or 1107  
CC AAs, domain 1,2, and partial domain 3-pref. where the protein is  
CC truncated at AA posn. 243; AAsn-271 is Asp and/or AAsn-300 is Asp; or the  
CC AA sequence without the N-linked glycosylation sites; of CD4 protein.  
CC Such fragments can also contain the corresp. AA of murine CD4 subseqn. at  
CC more than one triplet site in the DNA, selected from a triplet encoding  
CC the AA at posn. 48; 50; 51; 121; 122; 123; 155; 156; or 158 of human CD4.  
CC CDA fragments have the capacity to prevent infection of human T-  
CC lymphocyte infection by HIV and to prevent the formation of human T-  
CC lymphocyte syncytia. They can be used in the diagnosis, treatment and  
CC prevention of HIV infection. (Updated on 25-MAR-2003 to correct PR  
CC field.) (Updated on 25-MAR-2003 to correct PI field.)  
XX  
SQ Sequence 370 AA:  
Query Match 55.4%; Score 1891; DB 1; Length 370;  
Best Local Similarity 99.7%; Pred. No. 2.7e-94;  
Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 27 KVVIGKKGDTVELTCTASQKKSIOFHWKNSNQIKLGNQGSFLTKGPSKLNDRADSRSL 86  
DB 2 KVVIGKKGDTVELTCTASQKKSIOFHWKNSNQIKLGNQGSFLTKGPSKLNDRADSRSL 61  
QY 87 WDOGNFPLIITKNIKIEDSDTYICEVEDQKEVOLLVFGLTANSPTHTLLOQGSLLTLTLESP 146  
DB 62 WDOGNFPLIITKNIKIEDSDTYICEVEDQKEVOLLVFGLTANSPTHTLLOQGSLLTLTLESP 121  
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOKQKVEFKIDIVLAFOK 206  
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOKQKVEFKIDIVLAFOK 181  
QY 207 ASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMWQAERASSSKSWITFDLKNKEVSARKV 266  
DB 182 ASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMWQAERASSSKSWITFDLKNKEVSARKV 241  
QY 267 TODPKLQMGKLLPHLTLPOALPOYASGNLTLLALEKTHGLHDEVNLVYMRARQLOK 326  
DB 242 TODPKLQMGKLLPHLTLPOALPOYASGNLTLLALEKTHGLHDEVNLVYMRARQLOK 301  
QY 327 TCEVWGPSTPKMLSLKLENKAKVSKREKPVVYLNPEAGMOCCLLSDSGVLLSNIKY 386  
DB 302 TCEVWGPSTPKMLSLKLENKAKVSKREKPVVYLNPEAGMOCCLLSDSGVLLSNIKY 361  
QY 387 LPTWSTPV 394  
DB 362 LPTWSTPV 369

RESULT 98  
AAB83356  
ID AAB83356 standard; protein; 370 AA.  
XX  
XX AAB83356;  
AC  
XX  
DT 09-OCT-2001 (first entry)  
XX  
DE Human CD4 protein sequence.  
XX  
KM Chemotactic chemokine receptor 5; gp120; CD4; therapy; HIV; CCR5;  
KM human immunodeficiency virus; anti-inflammatory disease; human.  
XX  
OS Homo sapiens.  
XX  
PN EPI118858-A2.  
XX  
PD 25-JUL-2001.  
XX  
PF 03-JAN-2001; 2001EP-00300020.  
XX  
PR 12-JAN-2000; 2000GB-00000659.  
PR 12-JAN-2000; 2000GB-00000661.  
PR 12-JAN-2000; 2000GB-00000663.  
XX  
PA (PFIZ ) PFIZER LTD.  
PA (PFIZ ) PFIZER INC.  
PI Dobbs S, Petros M, Rickett GA;  
XX  
XX WPI; 2001-477088/52.  
DR N-PSDB; AAF87101.  
DR  
XX  
XX Determining if an agent can modulate CCR5-gp120 interaction, comprises  
PT incubating the agent with CCR5 and gp120 and determining if the agent  
PT modulates the interaction.  
XX  
XX Claim 1; Page 112; 113pp; English.  
PS  
XX  
XX This sequence represents the human CD4 protein sequence. The invention  
CC relates to a method for determining whether an agent is capable of  
CC modulating the interaction of chemotactic chemokine receptor 5 (CCR5)  
CC with gp120, comprising incubating the agent with CCR5 and gp120 and  
CC determining whether the agent modulates the interaction, where gp120 is  
CC associated with CD4, and where the interaction is a low affinity binding.  
CC The method is used to identify an agent capable of modulating the  
CC interaction of CCR5 with gp120. An agent identified by the method is used  
CC to prepare a pharmaceutical composition for the treatment of a disease or  
CC condition associated with CCR5 and gp120 interaction, to treat a subject  
CC with a disease or condition associated with CCR5 and gp120 interaction,  
CC and for preparing a pharmaceutical for treating human immunodeficiency  
CC virus (HIV). It can also be used to treat anti-inflammatory diseases. The method  
CC is commercially useful, amenable to high throughput screening, and  
CC detects interaction of gp120 with cells expressing only CCR5  
XX  
SQ Sequence 370 AA:  
Query Match 55.4%; Score 1891; DB 4; Length 370;  
Best Local Similarity 99.7%; Pred. No. 2.7e-94;  
Matches 367; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 27 KVVIGKKGDTVELTCTASQKKSIOFHWKNSNQIKLGNQGSFLTKGPSKLNDRADSRSL 86  
DB 2 KVVIGKKGDTVELTCTASQKKSIOFHWKNSNQIKLGNQGSFLTKGPSKLNDRADSRSL 61  
QY 87 WDOGNFPLIITKNIKIEDSDTYICEVEDQKEVOLLVFGLTANSPTHTLLOQGSLLTLTLESP 146  
DB 62 WDOGNFPLIITKNIKIEDSDTYICEVEDQKEVOLLVFGLTANSPTHTLLOQGSLLTLTLESP 121  
QY 147 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOKQKVEFKIDIVLAFOK 206  
DB 122 PGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWCTVLOKQKVEFKIDIVLAFOK 181

QY	20	ASSIVYKKKEQVEPSEFPPLAFVYVTKLTGSGELIMQAEASSSSKMTITDLNKKVSVKRV	265
Db	182	ASSIVYKKKEEQVEPSEFPPLAFVYVTKLTGSGELIMQAEASSSSKMTITDLNKKVSVKRV	241
QY	267	TOPPKLQMGMKKPLPHLHTLTPQALPQVAGSNNLTALAEATGKLHGEVNLVWRATOLQKNL	328
Db	242	TOPPKLQMGMKKPLPHLHTLTPQALPQVAGSNNLTALAEATGKLHGEVNLVWRATOLQKNL	301
QY	327	TCEVWGPTSPKMLSLIKLENKEAKVSKREKPVVVLNPEAGWMOCLSDSGVLLIESNIKV	386
Db	302	TCEVWGPTSPKMLSLIKLENKEAKVSKREKAVVVLNPEAGWMOCLSDSGVLLIESNIKV	361
QY	387	LPTWSTPV 394	
Db	362	LPTWSTPV 369	
RESULT 99			
ID	AAV54500	standard; protein; 433 AA.	
AC	AAV54500;		
XX			
XX	25-APR-2000	(first entry)	
XX			
DE		Amino acid sequence of the human CD4 protein.	
XX			
KW		Human; CD4 protein; antigenic peptide; CDR2-like domain; apoptosis;	
KW		syncytia formation; human immune deficiency virus; HIV binding;	
KW		CD4-Class II interaction; immunisation; CD4 surface complex;	
KW		immune response; transplant rejection; autoimmune disease;	
KW		rheumatoid arthritis; systemic lupus erythematosus; psoriasis.	
XX			
OS		Homo sapiens.	
XX			
FH	Key	Location/Qualifiers	
FT	Domain	27..66	
FT		/note="CDR-2 like domain; specifically claimed in claim	
FT		1"	
XX			
PN		W09967294-A1.	
PD		29-DEC-1999.	
XX			
PF		21-JUN-1999; 99MO-US014030.	
XX			
PR		20-JUN-1998; 98US-00100409.	
XX			
PA		(UNBI-) UNITED BIOMEDICAL INC.	
XX			
PI		Wang CY;	
XX			
DR		WPI; 2000-160579/14.	
XX			
PT		New antigenic peptide from the CDR2 domain of CD4, for immunization	
PT		against e.g. human immune deficiency virus.	
XX			
PS		Claim 1; Page 70-71; 106pp; English.	
XX			
CC		The present sequence represents the human CD4 protein. The specification	
CC		describes antigenic peptides derived from the CDR2-like domain of CD4	
CC		(amino acids 27-66 of AAV54500). These antigenic peptides present	
CC		neutralising receptor/co-receptor effector sites of the CDR2-like domain.	
CC		The peptides evoke effective antibody responses by having optimised site	
CC		specificity. The induced antibodies block human immune deficiency virus	
CC		(HIV) binding and syncytia formation. They may also block CD4-Class II	
CC		interactions with other cells, deliver signals to T cells (inhibiting	
CC		normal CD4+-mediated immunoregulatory functions) or induce apoptosis of	
CC		CD4 cells by simultaneous engagement of T cell receptors. Conjugates and	
CC		peptides containing the antigenic peptides are used for active	
CC		immunisation to generate antibodies against CD4 surface complexes,	
CC		especially to prevent binding of HIV to CD4 and thus HIV infection, but	

Query Match	Best Local Similarity	55.3% Matches 366; Conservative 0; Mismatches 2; Indels 0; Gaps 0	Score 1887; DB 3; Length 433; Pred. No. 5.2e-94;
CC	also to treat undesirable immune responses such as transplant rejection,		
CC	or autoimmune diseases (rheumatoid arthritis, systemic lupus		
CC	erythematosus or psoriasis). These conjugates produce high-titre		
CC	antibodies which are broadly neutralising against primary isolates from		
CC	all classes of HIV-1 and of HIV-2. The peptides may be cyclically		
CC	constrained and may include a promiscuous T helper epitope that is active		
CC	in genetically diverse subjects		
XX			
XX	Sequence 433 AA;		
QY	27 KVLVGGKKGDVELTCTASQKKSIOFHMKNNSNOIKILGNQGSFLLTKGPSKLNDRADSRSL	86	
Db	2 KVLVGGKKGDVELTCTASQKKSIOFHMKNNSNOIKILGNQGSFLLTKGPSKLNDRADSRSL	61	
QY	87 WDOGNFPLITIKNLKIEDSTTYICEVEDQEEVQLLVFGITANSDFHLLQGSLTITLESF	144	
Db	62 WDOGNFPLITIKNLKIEDSDTYICEVEDQEEVQLLVFGITANSDFHLLQGSLTITLESF	122	
QY	147 PGSSPSVQCRSPGKNIQGGKTLVSQLELQDSGTMTCTVLONOKKVEKIDIVLAFQK	200	
Db	122 PGSSPSVQCRSPGKNIQGGKTLVSQLELQDSGTMTCTVLONOKKVEKIDIVLAFQK	187	
QY	207 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELMWAQBRASSKSNITFDLNKKEVSXRV	266	
Db	182 ASSIVYKKEGEQVEFSFPLAFTVEKLTGSGELMWAQBRASSKSNITFDLNKKEVSXRV	241	
QY	267 TQDPKIQMGKKLPLHLITLPOALPQVAGSNLTLEAKTGKIHQEVNLVVMRATOLQKRL	322	
Db	242 TQDPKIQMGKKLPLHLITLPOALPQVAGSNLTLEAKTGKIHQEVNLVVMRATOLQKRL	307	
QY	327 TCEWVGTSFKMLSLKLENKAKVSKREKPYWVNLNPRAGMWCLLSPSGOVLLESNIKV	384	
Db	302 TCEWVGTSFKMLSLKLENKAKVSKREKPYWVNLNPRAGMWCLLSPSGOVLLESNIKV	367	
QY	387 LPTWSTPV 394		
Db	362 LPTWSTPV 369		
RESULT 100			
AA74222	AA74222 standard; protein; 432 AA.		
XX	AA74222;		
AC	25-MAR-2003 (revised)		
DT	26-NOV-1995 (first entry)		
XX	Epitope on the primary CD4 sequence.		
DE	Chimaeric; mutant; mapping; immunodiagnosics.		
KW	Synthetic.		
XX	OS		
XX	US5411861-A.		
PN	02-MAY-1995.		
XX	27-FEB-1992; 92US-00842465.		
XX	15-APR-1988; 88US-00181826.		
XX	(GEHO ) GEN HOSPITAL CORP.		
PA	Seed B, Peterson A;		
XX	WPI; 1995-178122/23.		
DR	Mutational analysis method for protein epitope(s) - by expressing mutant		
PT			

PT cDNA and using negative and positive selection to identify binding loss  
PT mutants.

PS Disclosure: Fig 2; 28pp; English.

XX The sequence is that of an epitope on the primary CD4 sequence. The  
CC epitope sequence can be identified by vector pIH3M. The See also  
CC AAR74221. (Updated on 25-MAR-2003 to correct PF field.)  
XX

SO Sequence 432 AA;

Query Match 55.2%; Score 1884; DB 2; Length 432;

Best Local Similarity 99.5%; Pred. No. 7.5e-94;

Matches 366; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY	27	KVLGKKGDTVELTCTAOKKSIOPHMKNNSQIKILGNQGSFLTKGPSKLANDRADSRSL	86
DB	1	KVLGKKGDTVELTCTAOKKSIOPHMKNNSQIKILGNQGSFLTKGPSKLANDRADSRSL	60
QY	87	WDOGNFPLIINKLKIEDSDTYICEVEDQKEEVOLLVFGLTANSPTHLQGSLLTLESF	146
DB	61	WDOGNFPLIINKLKIEDSDTYICEVEDQKEEVOLLVFGLTANSPTHLQGSLLTLESF	120
QY	147	PGSSPSVQCRSPRGKNIQSGKTLVSQLELQDSGTCTVLQNKVFEKIDIVVLAFOK	206
DB	121	PGSSPSVQCRSPRGKNIQSGKTLVSQLELQDSGTCTVLQNKVFEKIDIVVLAFOK	180
QY	207	ASSIVYKKEGEQVESPFLATVEKLTGSGELMWAERASSKSWITFDLKNKEVSQKV	266
DB	181	ASSIVYKKEGEQVESPFLATVEKLTGSGELMWAERASSKSWITFDLKNKEVSQKV	240
QY	267	TODPLOWGKKLPLHLTLPOALPOYAGSGNLTLEAKTGKLEHVNLVVMRATOLOKNT	326
DB	241	TODPLOWGKKLPLHLTLPOALPOYAGSGNLTLEAKTGKLEHVNLVVMRATOLOKNT	300
QY	327	TCEVWGPTSPKLMLSLKLENKAKVKREKPVWVLNPEAGMOCILSDSGVLLSNIKV	386
DB	301	TCEVWGPTSPKLMLSLKLENKAKVKREKPVWVLNPEAGMOCILSDSGVLLSNIKV	360
QY	387	LPTWSTPV 394	
DB	361	LPTWSTPV 368	

Search completed: August 3, 2004, 13:08:29  
Job time : 74.926 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:51:44 ; Search time 44.4955 Seconds  
(without alignments)  
4594.975 Million cell updates/sec

Title: SEQ7  
Perfect score: 3414  
Sequence: 1 MNRGVPRHLLVLQALLP.....DETCAGQDELDELMTTDP 648

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 125 summaries

Database :

SPTRMBL\_25:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_plant:\*  
10: sp\_rodent:\*  
11: sp\_virus:\*  
12: sp\_vertebrate:\*  
13: sp\_unclassified:\*  
14: sp\_virus:\*  
15: sp\_bacteria:\*  
16: sp\_bacteriap:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1705	49.9	009261	009261 cercopithec
2	1701	49.8	009260	009260 cercopithec
3	1700	49.8	095NE9	095NE9 cercopithec
4	1696	49.7	009259	009259 cercopithec
5	1696	49.7	009263	009263 cercopithec
6	1689	49.5	009262	009262 cercopithec
7	1688	49.4	002805	002805 cercopithec
8	1601.5	46.9	08H2T8	08H2T8 callithrix
9	1575.5	46.1	08H2T7	08H2T7 saimiri sci
10	1257	36.8	08NF17	08NF17 homo sapien
11	1221	35.8	07Z5M1	07Z5M1 homo sapien
12	1214	35.6	096P08	096P08 homo sapien
13	1211.5	35.5	07Z7P5	07Z7P5 homo sapien
14	1208.5	35.4	07Z351	07Z351 homo sapien
15	1187	34.8	09X578	09X578 delphinapte
16	1163	34.1	086TT2	086TT2 homo sapien

17	1160	34.0	521	4	08N4Y9	08N4Y9 homo sapien
18	1130	33.1	473	4	08TC63	08TC63 homo sapien
19	1087	31.8	474	6	P79355	P79355 felis silve
20	985.5	28.9	457	11	061396	061396 mus musculu
21	958	28.1	433	11	055054	055054 mus musculu
22	889	26.0	337	6	095M34	095M34 equus cabal
23	854	25.0	470	11	07TWT1	07TWT1 mus musculu
24	841.5	24.6	437	11	09RIA4	09RIA4 mus musculu
25	841.5	24.6	469	11	08R3V9	08R3V9 mus musculu
26	835	24.5	463	11	099LC4	099LC4 mus musculu
27	819	24.0	468	11	099LC1	099LC1 mus musculu
28	805.5	23.6	473	11	09DBL4	09DBL4 mus musculu
29	797.5	23.4	473	11	099125	099125 mus musculu
30	795.5	23.3	473	11	091205	091205 mus musculu
31	794.5	23.3	474	11	08R3H6	08R3H6 mus musculu
32	416.5	12.2	613	4	096E10	096E10 mus musculu
33	406.5	11.9	613	4	08WUX1	08WUX1 mus musculu
34	400	11.7	86	6	075596	075596 mandrillus
35	400	11.7	86	6	075597	075597 mandrillus
36	400	11.7	614	11	07TWT6	07TWT6 mus musculu
37	397	11.6	86	6	075594	075594 cercopithec
38	397	11.6	86	6	075593	075593 cercopithec
39	395	11.6	86	6	075595	075595 cercopithec
40	392	11.5	86	6	075598	075598 papio sp. (
41	391.5	11.5	597	4	096BB9	096BB9 homo sapien
42	388	11.4	86	6	077601	077601 lophocobus
43	388	11.4	614	4	096C86	096C86 homo sapien
44	383	11.2	86	6	077600	077600 lophocobus
45	382	11.2	618	4	096AA6	096AA6 homo sapien
46	380.5	11.1	613	11	08VCX7	08VCX7 mus musculu
47	377.5	11.1	597	4	09BU10	09BU10 homo sapien
48	375.5	11.0	597	4	09BOB8	09BOB8 homo sapien
49	372.5	10.9	588	4	08MUT4	08MUT4 homo sapien
50	357	10.5	71	4	013569	013569 mus musculu
51	356.5	10.4	375	4	086T11	086T11 homo sapien
52	346.5	9.6	684	13	090544	090544 ginglymosto
53	324.5	9.5	587	13	07TOR1	07TOR1 xenopus lae
54	313.5	9.2	478	4	07Z379	07Z379 homo sapien
55	305.5	8.9	99	6	029027	029027 sus scrofa
56	302	8.8	416	4	09NPE6	09NPE6 homo sapien
57	300.5	8.8	494	4	096KX8	096KX8 homo sapien
58	298	8.7	496	4	096KX8	096KX8 homo sapien
59	297	8.7	492	4	07Z374	07Z374 homo sapien
60	296	8.7	499	4	08NSK4	08NSK4 homo sapien
61	294	8.6	493	4	08NCL6	08NCL6 homo sapien
62	289	8.5	500	4	09BRV0	09BRV0 homo sapien
63	288.5	8.5	496	4	096DK0	096DK0 homo sapien
64	288	8.4	488	11	091WR1	091WR1 mus musculu
65	282.5	8.3	384	4	09UP60	09UP60 homo sapien
66	282.5	8.3	487	11	080217	080217 mus musculu
67	282.5	8.3	497	4	08W124	08W124 mus musculu
68	280.5	8.2	99	6	029028	029028 sus scrofa
69	278.5	8.2	489	11	08VCX4	08VCX4 mus musculu
70	276	8.1	481	11	091WT1	091WT1 mus musculu
71	276	8.1	482	11	08K172	08K172 mus musculu
72	274.5	8.0	487	11	099KX4	099KX4 mus musculu
73	274	8.0	484	11	08VEA0	08VEA0 mus musculu
74	272	8.0	481	11	091WT3	091WT3 mus musculu
75	272	8.0	482	11	091X92	091X92 mus musculu
76	271.5	8.0	479	11	091WPS	091WPS mus musculu
77	271.5	8.0	482	13	090MB5	090MB5 anas platyr
78	271.5	8.0	484	11	099LAE	099LAE mus musculu
79	271	7.9	120	11	088650	088650 marmota mon
80	270.5	7.9	486	11	091207	091207 mus musculu
81	270	7.9	488	11	08OCF2	08OCF2 mus musculu
82	269.5	7.9	426	11	09DCD9	09DCD9 mus musculu
83	269.5	7.9	480	11	091XEL	091XEL mus musculu
84	259.5	7.6	481	11	08ICV5	08ICV5 mus musculu
85	255	7.5	479	11	07TWT4	07TWT4 mus musculu
86	254.5	7.5	480	11	08K0Z4	08K0Z4 mus musculu
87	253.5	7.4	479	11	099M22	099M22 mus musculu
88	240	7.0	487	13	096V7	096V7 gallus gall
89	237	6.9	90	11	P70443	P70443 mus musculu

```

90 198 5 8 5636 4 096RW7 096RW7 homo sapien
91 186 5 4 258 13 090524 090524 ginglymosto
92 185.5 5 4 573 4 08WU38 08WU38 homo sapien
93 183.5 5 4 130 11 09DBW4 09DBW4 mus musculu
94 183.5 5 4 234 4 08N355 08N355 homo sapien
95 183.5 5 4 225 11 099M11 099M11 mus musculu
96 182.5 5 3 6620 4 096AA2 096AA2 homo sapien
97 179.5 5 3 2693 5 08ISF3 08ISF3 caenorhabd
98 179.5 5 3 2708 5 08ISF4 08ISF4 caenorhabd
99 179.5 5 3 2780 5 08MNS0 08MNS0 caenorhabd
100 179.5 5 3 18519 5 08ISF6 08ISF6 caenorhabd
101 179.5 5 3 18519 5 08ISF6 08ISF6 caenorhabd
102 179.5 5 2 18534 5 08ISF7 08ISF7 caenorhabd
103 178 5 2 17352 5 095YM2 095YM2 canis fami
104 178 5 2 17352 5 095YM2 095YM2 procambus
105 176.5 5 2 233 4 096T69 096T69 homo sapien
106 176.5 5 2 259 13 090530 090530 ginglymosto
107 175.5 5 1 1323 13 008476 008476 gallus gall
108 174.5 5 1 723 4 08N5F4 08N5F4 gallus gall
109 174.5 5 1 723 4 063669 063669 rattus norv
110 174.5 5 1 2673 4 096SC3 096SC3 homo sapien
111 174 5 1 236 4 096S61 096S61 homo sapien
112 174 5 1 8081 5 072120 072120 caenorhabd
113 172.5 5 1 234 4 0722U7 0722U7 homo sapien
114 172.5 5 1 4162 13 098918 098918 gallus gall
115 172.5 5 1 7962 4 010465 010465 homo sapien
116 172.5 5 1 34350 4 08WZ42 08WZ42 homo sapien
117 172 5 1 233 4 08TBC9 08TBC9 homo sapien
118 172 5 0 1031 13 090YM2 090YM2 brachydanto
119 171.5 5 0 257 13 090536 090536 ginglymosto
120 171 5 0 650 6 09GKR2 09GKR2 bos taurus
121 171 5 0 739 6 09GKR3 09GKR3 bos taurus
122 170 5 0 237 13 090545 090545 ginglymosto
123 170 5 0 238 7 09MXA2 09MXA2 aulonocara
124 170 5 0 379 11 08BLX5 08BLX5 mus musculu
125 170 5 0 397 11 08BFX8 08BFX8 mus musculu

```

## ALIGNMENTS

```

RESULT 1
009261 PRELIMINARY; PRT; 397 AA.
AC 009261;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 (Fragment) .
GN CD4.
OS Cercopithecus sabaues.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_Taxid=60711;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RA "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001225; AAB60872.1; -.
DR HSSP; P01730; IMIO.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TCS6.
DR InterPro; IPR007110; IG_Like.
DR Pfam; PF00047; IG_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IG; 1.
DR SMART; SM00406; IG; 1.

```

```

DR PROSITE; PSS0835; IG_Like; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43945 MW; 80C963B92A868CD3 CRC64;
Query Match
Best local similarity 89.9%; Pred. No. 1.4e-122;
Matches 330; Conservative 15; Mismatches 22; Indels 0; Gaps 0;
QY 28 VILGKGDVLELTCTASQKKSIOFHWNKSNQIKIIGNQSPFLTKGPKLNDRAISRSLW 87
1 VILGKGDVLELTCTASQKKSIOFHWNKSNQIKIIGNQSPFLTKGPKLNDRAISRSLW 60
DB DOGNEFLIKNKIDSDPTVCEVEDOKREVOLLVGLRANSDTHLLOGQSITLTRESP 147
61 DGQCSMITKNIKIDSETVYICEVENKKEVELVFGLRANSDTHLLOGQSITLTRESP 120
QY 148 GSSPSVQCRSPRGNKNIQGGKTLVSQLELDSDGTCTVLQNKVPEKIDIVLAFQKA 207
121 GSSPSVQCRSPRGNKNIQGGKTLVSQLELDSDGTCTVLQNKVPEKIDIVLAFQKA 180
DB 208 SSIYKKGEGVEFSPPLAFVYEKLTGSGELMWQBRASSSGWTFDLKNKEVYKRYT 267
181 SSTYKKGEGVEFSPPLAFVYEKLTGSGELMWQBRASSSGWTFDLKNKEVYKRYT 240
QY 268 ODPKLOMGKKLPLHITLPALPOYAGSGVLTALFAKTKLHOENVLYVMRATOLQKNLT 327
241 ODPKLOMGKKLPLHITLPALPOYAGSGVLTALFAKTKLHOENVLYVMRATOLQKNLT 300
QY 328 CEWVGPTSPKMLSLKENKEAKVSRKRPVAVNLPEAGMOCLLSDSGOVLLSNIKYL 387
301 CEWVGPTSPKMLSLKENKEAKVSRKRPVAVNLPEAGMOCLLSDSGOVLLSNIKYL 360
DB 368 PTWSTPV 394
361 PTWSTPV 367
DB

```

```

RESULT 2
009260 PRELIMINARY; PRT; 397 AA.
AC 009260;
DT 01-JUL-1997 (TREMBlrel. 04, Created)
DT 01-JUL-1997 (TREMBlrel. 04, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE CD4 (Fragment) .
GN CD4.
OS Cercopithecus sabaues.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCBI_Taxid=60711;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RA "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001224; AAB60871.1; -.
DR HSSP; P01730; IMIO.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TCS6.
DR InterPro; IPR007110; IG_Like.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IG; 1.
DR PROSITE; PSS0835; IG_Like; 1.
FT NON_TER 1
FT NON_TER 397

```

Seq: 397 AA; 43882 MW; 4788277E992EEB9 CRC64;  
Query Match 49.8%; Score 1701; DB 6; Length 397;  
Best Local Similarity 89.6%; Pred. No. 2.8e-122;  
Matches 329; Conservative 16; Mismatches 22; Indels 0; Gaps 0;  
QY 28 VVLGKKGDVVELTCTASQKKSIOFHWNQSIKILGNQSFLLTGPSKLNDRADSRSLW 87  
1 VVLGKKGDVVELTCTASQKKSIOFHWNQSIKILGNQSFLLTGPSKLNDRADSRSLW 60  
DB 1 VVLGKKGDVVELTCTASQKKSIOFHWNQSIKILGNQSFLLTGPSKLNDRADSRSLW 60  
QY 88 DQGNFLIKNLKIEBDDTYICEVEDQEVQLVFLTANSPDTHLLOQGSLLTLESPP 147  
61 DQGNFLIKNLKIEBDDTYICEVEDQEVQLVFLTANSPDTHLLOQGSLLTLESPP 120  
DB 61 DQGNFLIKNLKIEBDDTYICEVEDQEVQLVFLTANSPDTHLLOQGSLLTLESPP 120  
QY 148 GSSPSVQCSRPGRKNIQGGKTLISVQLBLQDSGTWTCTVLOKQKVEFKIDIVLAFOKA 207  
121 GSSPSVQCSRPGRKNIQGGKTLISVQLBLQDSGTWTCTVLOKQKVEFKIDIVLAFOKA 180  
DB 121 GSSPSVQCSRPGRKNIQGGKTLISVQLBLQDSGTWTCTVLOKQKVEFKIDIVLAFOKA 180  
QY 208 SSIYKKEGQVFEFPLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSVKRYT 267  
181 SSIYKKEGQVFEFPLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSVKRYT 240  
DB 181 SSIYKKEGQVFEFPLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSVKRYT 240  
QY 268 QDPKLGKGLPMLTLPLPOLPYAGSGNLTALBAKTKLHOEVNLVWRATOLQKNT 327  
241 QDPKLGKGLPMLTLPLPOLPYAGSGNLTALBAKTKLHOEVNLVWRATOLQKNT 300  
DB 241 QDPKLGKGLPMLTLPLPOLPYAGSGNLTALBAKTKLHOEVNLVWRATOLQKNT 300  
QY 328 CEVWGPTSPKMLSLKLENKAATVSKREKPVWVLPNPAAGMOCCLSDSGVLLSNIKVL 387  
301 CEVWGPTSPKMLSLKLENKAATVSKREKPVWVLPNPAAGMOCCLSDSGVLLSNIKVL 360  
DB 301 CEVWGPTSPKMLSLKLENKAATVSKREKPVWVLPNPAAGMOCCLSDSGVLLSNIKVL 360  
QY 388 PTWSTPV 394  
361 PTWSTPV 367  
DB 361 PTWSTPV 367

RESULT 3  
Q95NE9 PRELIMINARY; PRT; 397 AA.  
ID Q95NE9  
AC Q95NE9  
DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CD4 (Fragment).  
GN GN  
OS Cercopithecus pygerythrus.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
OC Cercopithecinae; Cercopithecus.  
OX NCBI\_TaxID=60710;  
RN RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98017879; PubMed=9379478;  
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,  
Corbet S., Barre-Sinoussi F., Allan J.S.,  
"Relation between phylogeny of African green monkey CD4 genes and  
their respective simian immunodeficiency virus genes."  
RT J. Med. Primatol. 26:120-128(1997).  
RL J. Med. Primatol. 26:120-128(1997).  
DR EMBL; AF001227; AAB60874.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0006955; P:immune response; IEA.  
DR InterPro; IPR000973; CD4\_TcAg.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig\_2.  
DR PRINTS; PRO0692; CD4TCANTIGEN.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PSS0835; IG\_LIKE; 1.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 397 AA; 43946 MW; 21C3530882ABFC0 CRC64;  
Query Match 49.8%; Score 1700; DB 6; Length 397;  
Best Local Similarity 89.6%; Pred. No. 3.3e-122;

Matches 329; Conservative 15; Mismatches 23; Indels 0; Gaps 0;  
QY 28 VVLGKKGDVVELTCTASQKKSIOFHWNQSIKILGNQSFLLTGPSKLNDRADSRSLW 87  
1 VVLGKKGDVVELTCTASQKKSIOFHWNQSIKILGNQSFLLTGPSKLNDRADSRSLW 60  
DB 1 VVLGKKGDVVELTCTASQKKSIOFHWNQSIKILGNQSFLLTGPSKLNDRADSRSLW 60  
QY 88 DQGNFLIKNLKIEBDDTYICEVEDQEVQLVFLTANSPDTHLLOQGSLLTLESPP 147  
61 DQGNFLIKNLKIEBDDTYICEVEDQEVQLVFLTANSPDTHLLOQGSLLTLESPP 120  
DB 61 DQGNFLIKNLKIEBDDTYICEVEDQEVQLVFLTANSPDTHLLOQGSLLTLESPP 120  
QY 148 GSSPSVQCSRPGRKNIQGGKTLISVQLBLQDSGTWTCTVLOKQKVEFKIDIVLAFOKA 207  
121 GSSPSVQCSRPGRKNIQGGKTLISVQLBLQDSGTWTCTVLOKQKVEFKIDIVLAFOKA 180  
DB 121 GSSPSVQCSRPGRKNIQGGKTLISVQLBLQDSGTWTCTVLOKQKVEFKIDIVLAFOKA 180  
QY 208 SSIYKKEGQVFEFPLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSVKRYT 267  
181 SSIYKKEGQVFEFPLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSVKRYT 240  
DB 181 SSIYKKEGQVFEFPLAFTVEKLTGSGELMWAERASSKSWITFDLNKKEVSVKRYT 240  
QY 268 QDPKLGKGLPMLTLPLPOLPYAGSGNLTALBAKTKLHOEVNLVWRATOLQKNT 327  
241 QDPKLGKGLPMLTLPLPOLPYAGSGNLTALBAKTKLHOEVNLVWRATOLQKNT 300  
DB 241 QDPKLGKGLPMLTLPLPOLPYAGSGNLTALBAKTKLHOEVNLVWRATOLQKNT 300  
QY 328 CEVWGPTSPKMLSLKLENKAATVSKREKPVWVLPNPAAGMOCCLSDSGVLLSNIKVL 387  
301 CEVWGPTSPKMLSLKLENKAATVSKREKPVWVLPNPAAGMOCCLSDSGVLLSNIKVL 360  
DB 301 CEVWGPTSPKMLSLKLENKAATVSKREKPVWVLPNPAAGMOCCLSDSGVLLSNIKVL 360  
QY 388 PTWSTPV 394  
361 PTWSTPV 367  
DB 361 PTWSTPV 367

RESULT 4  
O09259 PRELIMINARY; PRT; 397 AA.  
ID O09259  
AC O09259  
DT 01-JUL-1997 (TrEMBLrel. 04, Created)  
DT 01-JUL-1997 (TrEMBLrel. 04, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE CD4 (Fragment).  
GN GN  
OS Cercopithecus aethiops.  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;  
OC Cercopithecinae; Cercopithecus.  
OX NCBI\_TaxID=60711;  
RN RN  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98017879; PubMed=9379478;  
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,  
Corbet S., Barre-Sinoussi F., Allan J.S.,  
"Relation between phylogeny of African green monkey CD4 genes and  
their respective simian immunodeficiency virus genes."  
RT J. Med. Primatol. 26:120-128(1997).  
RL J. Med. Primatol. 26:120-128(1997).  
DR EMBL; AF001223; AAB60870.1; -.  
DR HSSP; P01730; IMIO.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0006955; P:immune response; IEA.  
DR InterPro; IPR000973; CD4\_TcAg.  
DR InterPro; IPR003596; Ig\_LIKE.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig\_2.  
DR PRINTS; PRO0692; CD4TCANTIGEN.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PSS0835; IG\_LIKE; 1.  
FT NON\_TER 1  
FT NON\_TER 1  
SQ SEQUENCE 397 AA; 43881 MW; 7CE39AD0F8506C81 CRC64;  
Query Match 49.7%; Score 1696; DB 6; Length 397;  
Best Local Similarity 89.4%; Pred. No. 6.7e-122;  
Matches 328; Conservative 17; Mismatches 22; Indels 0; Gaps 0;  
QY 28 VVLGKKGDVVELTCTASQKKSIOFHWNQSIKILGNQSFLLTGPSKLNDRADSRSLW 87

```
Db 1 VVLGKKGDVVELTCNASQTTTQFHHKNSNQIKILGQSSFLTGGSSKLDRIDSRSLW 60
Qy 88 DQGNFPLIINKLKIEDSDTYICEVEDQKEVOLVFGLTANSPTHLLQGOSLTLTLESP 147
Db 61 DQGCFSMIIKNIKIEDSETTYICEVENKEBEVLVGLTANSPTHLLQGOSLTLTLESP 120
Qy 148 GSSPSVQCRSPRGKNIQGGKTLVSQLELODSGTWTCVLQONKVEFKIDIVLAFOKA 207
Db 121 GSSPSVKCRSPRGKNIQGGRTLSVPQLERODSGTWTCVSNQONVTEFKIDIVLAFOKA 180
Qy 208 SSTVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSSKSWITFDLKNKESVKRYT 267
Db 181 SSTVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSSKSWITFDLKNKESVKRYT 240
Qy 268 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHQEVNLVVMRATOLQKNLT 327
Db 241 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHQEVNLVVMRATOLQKNLT 300
Qy 328 CEVWGPTSPKMLSLKENKEAKVSKREKPVWVLNPEAGMOCLLSDSQVLLSNIKVL 387
Db 301 CEVWGPTSPKMLSLKENKAAVSKQAKAVWVLNPEAGMOCLLSDSQVLLSNIKVL 360
Qy 388 PTWSTPV 394
Db 361 PTWPTPV 367
```

## RESULT 5

```
009263 PRELIMINARY; PRT; 397 AA.
ID 009263;
AC 009263;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE CD4 (Fragment).
GN CD4.
OS Cercopithecus tantalus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCB1_TaxID=60712;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001221; AAB60869.1; -.
DR HSSP; P01730; IMIQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PSS0835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43994 MW; A3CD031535A51524 CRC64;
```

## Query Match 49.7%; Score 1696; DB 6; Length 397;

Best Local Similarity 89.4%; Pred. No. 6.7e-122;

Matches 328; Conservative 15; Mismatches 24; Indels 0; Gaps 0;

```
Qy 28 VVLGKKGDVVELTCNASQTTTQFHHKNSNQIKILGQSSFLTGGSSKLDRIDSRSLW 87
Db 1 VVLGKKGDVVELTCNASQTTTQFHHKNSNQIKILGQSSFLTGGSSKLDRIDSRSLW 60
```

```
Qy 88 DQGNFPLIINKLKIEDSDTYICEVEDQKEVOLVFGLTANSPTHLLQGOSLTLTLESP 147
Db 61 DQGCFSMIIKNIKIEDSETTYICEVENKEBEVLVGLTANSPTHLLQGOSLTLTLESP 120
Qy 148 GSSPSVQCRSPRGKNIQGGKTLVSQLELODSGTWTCVLQONKVEFKIDIVLAFOKA 207
Db 121 GSSPSVKCRSPRGKNIQGGRTLSVPQLERODSGTWTCVSNQONVTEFKIDIVLAFOKA 180
Qy 208 SSTVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSSKSWITFDLKNKESVKRYT 267
Db 181 SSTVYKKEGEQVEFSFPLAFVTEKLTGSGELMWQAEARASSSKSWITFDLKNKESVKRYT 240
Qy 268 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHQEVNLVVMRATOLQKNLT 327
Db 241 QDPKLOMGKKLPLHLTLPOALPOYAGSGNLTALBAKTKLHQEVNLVVMRATOLQKNLT 300
Qy 328 CEVWGPTSPKMLSLKENKEAKVSKREKPVWVLNPEAGMOCLLSDSQVLLSNIKVL 387
Db 301 CEVWGPTSPKMLSLKENKAAVSKQAKAVWVLNPEAGMOCLLSDSQVLLSNIKVL 360
Qy 388 PTWSTPV 394
Db 361 PTWPTPV 367
```

## RESULT 6

```
009262 PRELIMINARY; PRT; 397 AA.
ID 009262;
AC 009262;
DT 01-JUL-1997 (TReMBLrel. 04, Created)
DT 01-JUL-1997 (TReMBLrel. 04, Last sequence update)
DT 01-OCT-2003 (TReMBLrel. 25, Last annotation update)
DE CD4 (Fragment).
GN CD4.
OS Cercopithecus tantalus.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Cercopithecus.
OX NCB1_TaxID=60712;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Muller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
DR EMBL; AF001221; AAB60868.1; -.
DR HSSP; P01730; IMIQ.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0006955; P:immune response; IEA.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PSS0835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 397
SQ SEQUENCE 397 AA; 43954 MW; CF7F2F5D82335B0D CRC64;
```

## Query Match 49.5%; Score 1689; DB 6; Length 397;

Best Local Similarity 89.4%; Pred. No. 2.3e-121;

Matches 328; Conservative 15; Mismatches 24; Indels 0; Gaps 0;

```
Qy 28 VVLGKKGDVVELTCNASQTTTQFHHKNSNQIKILGQSSFLTGGSSKLDRIDSRSLW 87
Db 1 VVLGKKGDVVELTCNASQTTTQFHHKNSNQIKILGQSSFLTGGSSKLDRIDSRSLW 60
Qy 88 DQGNFPLIINKLKIEDSDTYICEVEDQKEVOLVFGLTANSPTHLLQGOSLTLTLESP 147
Db 61 DQGCFSMIIKNIKIEDSETTYICEVENKEBEVLVGLTANSPTHLLQGOSLTLTLESP 120
```



```

Qy 148 GSSPSVOCRRPRGNIOGKTLVSOLELSDSGTWTCTVLOKKEVEFKIDIVLAFOKA 207
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GSSPSVOCRRPRGNIOGKTLVSOLELSDSGTWTCTVLOKKEVEFKIDIVLAFOKA 180
Qy 208 SSIYKKEGEQVEFSFPLAFTVEKLTGSGELMWAERASSSSKSWITFDLKNKEVSVKRYT 267
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SSIYKKEGEQVEFSFPLAFTVEKLTGSGELMWAERASSSSKSWITFDLKNKEVSVKRYT 240
Qy 268 QDPKLQNGKKLPHLTLTPQALPOYAGSGNLTALAEATGKLGHOEVLVVMRATQLOKRLT 327
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 QDPKLQNGKKLPHLTLTPQALPOYAGSGNLTALAEATGKLGHOEVLVVMRATQLOKRLT 300
Qy 328 CEVWGPTSPKLMLSLKLENKAKYSKREKPYVWLNPEAGMOCILSDSGVLESNIKVL 387
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 CEVWGPTSPKLMLSLKLENKAKYTSKQAKAVWLNPEAGMOCILSDSGVLESNIKVL 360
Qy 388 PTWSTPV 394
    |||:|||||:
Db 361 PTWSTPV 367

```

## RESULT 7

```

ID 002805 PRELIMINARY; PRT; 397 AA.
AC 002805; 077593;
DT 01-JUL-1997 (Tremblrel. 04, Created)
DT 01-JUL-1997 (Tremblrel. 04, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)
    (Fragment).
GN CD4.
OS Cercopithecus aethiops (Green monkey) (Griwet).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9534;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Peripheral blood;
RA MEDLINE=98017879; PubMed=9379478;
RA Fomsgaard A., Møller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,
RA Corbet S., Barre-Sinoussi F., Allan J.S.;
RT "Relation between phylogeny of African green monkey CD4 genes and
RT their respective simian immunodeficiency virus genes.";
RL J. Med. Primatol. 26:120-128(1997).
RN [2]
RP SEQUENCE OF 80-165 FROM N.A.
RA MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (Primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC ONE V-LIKE AND 3 C2-LIKE DOMAINS.
DR EMBL: AF001226; AAB60873.1; -
DR EMBL: AF057380; AAC5124.1; -
DR HSP; P01730; IWIQ.
DR GO: GO:0016021; C:Integral to membrane; IEA.
DR GO: GO:0006955; P:Immune response; IEA.
DR InterPro: IPR000973; CD4_TGAG.
DR InterPro: IPR007110; IG-Like.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; 19; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV.1.
KM Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
    Lipoprotein; Palmitate; Repeat.

```

```

FT NON TER 1 1
FT DOMAIN <1 370
FT TRASNEM 371 391
FT DOMAIN 392 >397
FT DOMAIN <1 98
FT DOMAIN 99 176
FT DOMAIN 177 290
FT DOMAIN 291 347
FT CARBOHYD 15 15
FT CARBOHYD 30 30
FT CARBOHYD 269 269
FT CARBOHYD 298 298
FT DISULFID 14 82
FT DISULFID 128 157
FT DISULFID 301 343
FT LIPID 392 392
FT LIPID 395 395
FT NON TER 397 397
SQ SEQUENCE 397 AA; 43980 MW; F74C42E22B196155 CRC64;

```

Query Match 49.4%; Score 1688; DB 6; Length 397;  
 Best Local Similarity 89.1%; Pred. No. 2.8e-121;  
 Matches 327; Conservative 15; Mismatches 25; Indels 0; Gaps 0;

```

Qy 28 VILGKGDYELCTASQKSIQFMWNSNQIKLQSGSFLTKGPKLNDRADRSRLW 87
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VILGKGDYELCTASQKSIQFMWNSNQIKLQSGSFLTKGPKLNDRADRSRLW 60
Qy 88 DQGNFPLIINKLIEDSDYICVEVDQKEEVQLVFGLTNSDTHLQGSLLTLESPP 147
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 DQGNFPLIINKLIEDSDYICVEVDQKEEVQLVFGLTNSDTHLQGSLLTLESPP 120
Qy 148 GSSPSVOCRRPRGNIOGKTLVSOLELSDSGTWTCTVLOKKEVEFKIDIVLAFOKA 207
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 GSSPSVOCRRPRGNIOGKTLVSOLELSDSGTWTCTVLOKKEVEFKIDIVLAFOKA 180
Qy 208 SSIYKKEGEQVEFSFPLAFTVEKLTGSGELMWAERASSSSKSWITFDLKNKEVSVKRYT 267
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 SSIYKKEGEQVEFSFPLAFTVEKLTGSGELMWAERASSSSKSWITFDLKNKEVSVKRYT 240
Qy 268 QDPKLQNGKKLPHLTLTPQALPOYAGSGNLTALAEATGKLGHOEVLVVMRATQLOKRLT 327
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 QDPKLQNGKKLPHLTLTPQALPOYAGSGNLTALAEATGKLGHOEVLVVMRATQLOKRLT 300
Qy 328 CEVWGPTSPKLMLSLKLENKAKYSKREKPYVWLNPEAGMOCILSDSGVLESNIKVL 387
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 CEVWGPTSPKLMLSLKLENKAKYTSKQAKAVWLNPEAGMOCILSDSGVLESNIKVL 360
Qy 388 PTWSTPV 394
    |||:|||||:
Db 361 PTWSTPV 367

```

## RESULT 8

```

ID 08H278 PRELIMINARY; PRT; 457 AA.
AC 08H278;
DT 01-MAR-2003 (Tremblrel. 23, Created)
DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Lymphocyte antigen CD4.
OS Callithrix jacchus (Common marmoset).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Callitrichidae; Callitrich.
OX NCBI_TaxID=9483;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=22174698; PubMed=12186836;
RA Labonte J.A., Babcock G.J., Patel T., Sodroski J.;
RT "Blockade of HIV-1 Infection of New World Monkey Cells Occurs
RT Primarily at the Stage of Virus Entry.";
RL J. Exp. Med. 196:431-445(2002).
DR EMBL: AF452616; AAU14532.1; -

```



Matches	233	Conservative	11	Mismatches	11	Indels	6	Gaps	2
Qy	332	TPVPP-APEPSCCKTHTC----	PELLGSGSVLFPPEKDTLMISRTPEVTCVVVD	445					
Db	200	TPPPCRCPPEPKSCPTPPCCPCRCAPABELIGSGSVLPFPKPKDTLMSRTPEVTCVVVD	259						
Qy	446	SHEDEVFNNVVDVEVHNKTKRKREQYNSVTVSVLTLHODMNLNGEKYKCKVSNK	505						
Db	260	SHEDEVFKNVVDGEVHNNAKTKREDFNSTFRVSVLTLHODMNLNGEKYKCKVSNK	319						
Qy	506	ALPAPIEKTISKAKQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQ	565						
Db	320	ALPAPIEKTISKTKQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQ	379						
Qy	566	PENNYKTTTPVLVDSGGSFFLVSKLTVDSKRWQGSVFCSCVMHEALHNHYTQKSLSLSPG	625						
Db	380	PENNYNTTPPMLDSGGSFFLVSKLTVDSKRWQGSVFCSCVMHEALHNHYTQKSLSLSPG	439						
Qy	626	LQDDECAADGDELGLMTT	646						
Db	440	LQLESCAEADGDELGLMTT	460						
RESULT 11									
ID	Q725W1	PRELIMINARY;	PRT;	470	AA.				
AC	Q725W1								
DT	01-OCT-2003	(TREMBLrel. 25, Created)							
DT	01-OCT-2003	(TREMBLrel. 25, Last sequence update)							
DE	01-OCT-2003	(TREMBLrel. 25, Last annotation update)							
DE	Hypothetical protein.								
OS	Homo sapiens (Human).								
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;								
OC	Mammalia; Euteria; Primates; Catarrhini; Homidae; Homo.								
OX	NCBI_Taxid=9606;								
NP	SEQUENCE FROM N.A.								
RC	TISSUE=Spleen;								
RX	MEDLINE=2238257; PubMed=12477932;								
RA	Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,								
RA	Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,								
RA	Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,								
RA	Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,								
RA	Ditschenko L., Marushina K., Farmer A.A., Rubin G.M., Hong L.,								
RA	Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,								
RA	Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Prange C.,								
RA	Raha S.S., Loguelfano N.A., Peters G.J., Abramson R.D., Mullany S.J.,								
RA	Bosak S.A., McKean P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,								
RA	Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,								
RA	Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,								
RA	Foley J., Heltón E., Kerteman M., Madan A., Rodrigues S., Sanchez A.,								
RA	Whiting R.W., Madan A., Young A.C., Shevchenko V., Bouffard G.G.,								
RA	Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,								
RA	Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,								
RA	Kryzhanovskii M.I., Skalska U., Smalov D.E., Schnerch A., Schein J.E.,								
RA	Jones S.J., Marz M.A.;								
RT	"Generation and initial analysis of more than 15,000 full-length human								
RT	and mouse cDNA sequences."								
RL	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).								
RP	SEQUENCE FROM N.A.								
RC	TISSUE=Spleen;								
RA	Strauberg R.;								
RL	Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.								
DR	EMBL; BC053984; AAH53984.1; -								
KM	Hypothetical protein.								
SO	SEQUENCE	470	AA;	778	CFJ4521483E1A	CRC64;			
Query Match									
		35.8%;	Score 1221;	DB 4;	Length 470;				
Best Local Similarity		46.4%;	Pred. No. 2,6e-85;						
Matches		284;	Conservative	36;	Mism				

Db	30	LVQPGSCIRLSVCVASGFTLANNYDMHWROGIGKLEWVNSKIGTAGDRRYVAGSVKGRFTIS	89
Qy	79	RADSRRLSDMGNPPLLIINKLIEDSDTYICEVDQKEEYVQLVFGLTANSDTLLQOSG	138
Db	90	RENKADSLYLQMN-----SLRVDAAVYIC-----ARAGMWAPLGARDINGCG	133
Qy	139	LTLTLESPGSSPSVQCKSPRGKNIQGGKTLVSQLELDQSGTWTCTVLQNKVKERID	198
Db	134	TMVTVSSASTGPFVFLPAPSSKSTSGG-TAALCCL-----	168
Qy	159	IVLVAFAQASSIYVKKGEQVEFEPFLAFYVEKLTGSGELMWAQERASSSKSWITPDLKN	258
Db	169	-----VKDYFPEPVTYS-----VMSGALTSG-----	189
Qy	259	KEVSIVKVTQDPKTLQMGKPLRHLTLQALPQVAGSGLTLALEAKTGKJLHOEVNLVVMR	318
Db	190	-----VH-TFPAVL-QSSGLXSLSSVTVTPBSSLSGTQTYI-----	222
Qy	319	ATOLQKMLTCEVWGPTSPKMLSLKENKEAKVSKREKPVWVLNPEAGMWOCCLSDSGOV	378
Db	223	-----CNV-----NHKSNKTKVDKV-----	238
Qy	379	LLESNIKLPFWSTPVPCCPAPEPSCDKTHTC-----PELLGGSVFLFPKPKDTLMIS	433
Db	239	-----EPKSCDKHTCPCCPAPELLGGSVFLFPKPKDTLMIS	277
Qy	434	RTPEVTCVVDVSHEDPEVKFNMTVVDGVEVNAATKREDOYNSYTVVSVTLTHODML	493
Db	278	RTPEVTCVVDVSHEDPEVKFNMTVVDGVEVNAATKREDOYNSYTVVSVTLTHODML	337
Qy	494	NGKEYCKVSNKALPAPIEKTISKAKQPREPOVYTLPSRDELTKNQVSLTCLVKGFYP	553
Db	338	NGKEYCKVSNKALPAPIEKTISKAKQPREPOVYTLPSRDELTKNQVSLTCLVKGFYP	397
Qy	554	SDIAYEWESNQPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQQGNVFSQSVNHEALHN	613
Db	398	SDIAYEWESNQPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQQGNVFSQSVNHEALHN	457
Qy	614	HYTOKSLSLSPG 625	
Db	458	HYTOKSLSLSPG 469	
RESULT 12			
Q96PQ8			
ID	Q96PQ8	PRELIMINARY;	PRT; 679 AA.
AC	Q96PQ8:	01-DEC-2001 (TREMBLrel. 19, Created)	
DT	01-JUN-2003 (TREMBLrel. 24, Last sequence update)		
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)		
DE	Factor VII active site mutant immunocjugate.		
OS	Homo sapiens (Human).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
RN	RP	SEQUENCE FROM N.A.	
RX	MEBLINE=21477448; PubMed=1159303;		
RA	Hu Z., Garen A.;		
RT	"Targeting tissue factor on tumor vascular endothelial cells and tumor		
RT	cells for immunotherapy in mouse models of prostatic cancer.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 98:12180-12185(2001).		
RP	SEQUENCE FROM N.A.		
RA	Hu Z., Garen A.;		
RL	Submitted (FEB-2003) to the EMBL/GenBank/DBJ databases.		
DR	EMBL; AF272774; AAKS8686.2;		
DR	GO; GO:0005576; C:extracellular binding; IEA.		
DR	GO; GO:0005509; F:calcium ion binding; IEA.		
DR	GO; GO:0004263; F:chymotrypsin activity; IEA.		
DR	GO; GO:0004295; F:trypsin activity; IEA.		
DR	GO; GO:0006508; P:proteolysis and peptidolysis; IEA.		



Seq	Sequence	482 AA;	52852 MM;	EDA75F1901D1A034	CNC64;
Query Match	35.4%	Score 1208.5;	DB 4;	Length 482;	
Best Local Similarity	97.4%;	Prod. No. 2.5e-84;			
Matches	225;	Conservative	0;	Mismatches 81;	Indels 5; Gaps 1;
QY	400	EPKSCDKHTC-----PELLGSPVFLPRPKDTLMISRTPEVTCVVVDVSHEDPEVK	454		
DB	251	EPKSCDKHTCPRCAPELLGSPVFLPRPKDTLMISRTPEVTCVVVDVSHEDPEVK	310		
QY	455	NWYVDGVEVHNAKTKRPREQYNSTRYVSVLTLVHOMDLNGEKYCKSNALPAPIKT	514		
DB	311	NWYVDGVEVHNAKTKRPREQYNSTRYVSVLTLVHOMDLNGEKYCKSNALPAPIKT	370		
QY	515	ISKAKGPREQVYTLPPSRDELTKQVSLTCLVGFYPSDIAVEMESNGQPENNYKTTP	574		
DB	371	ISKAKGPREQVYTLPPSRDELTKQVSLTCLVGFYPSDIAVEMESNGQPENNYKTTP	430		
QY	575	PVLDSGDFFLYSLKTLVDKSRWQGNVSCSVMEHALNNHTQKSLSLSPG	625		
DB	431	PVLDSGDFFLYSLKTLVDKSRWQGNVSCSVMEHALNNHTQKSLSLSPG	481		

	RESULT	15
ID	Q9X578	PRELIMINARY; PRT; 455 AA.
AC	Q9X578;	
DT	01-NOV-1999	(TREMBLrel. 12, Created)
DT	01-NOV-1999	(TREMBLrel. 12, Last sequence update)
DT	01-OCT-2003	(TREMBLrel. 25, last annotation update)
DE	T-cell surface glycoprotein CD4.	
OS	Delphinapterus leucas (Beluga whale).	
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Mammalia; Eutheria; Cetartiodactyla; Cetacea; Odontoceti;	
OX	NCBI_TaxId=9749;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RC	TISSUE=Thymus;	
RX	MEDLINE=99216435; PubMed=1099913;	
RA	Romano T.A., Ridgway S.H., Felten D.L., Quaranta V.;	
RT	"Molecular cloning and characterization of CD4 in an aquatic mammal,	
RL	the white whale Delphinapterus leucas.";	
DR	Immunogenetics 49:376-383(1999).	
DR	EMBL; AF0711799; AAD33738.1; -.	
DR	HSSP; P01730; 1WTO.	
DR	GO; GO:0016030; C:membrane; IEA.	
DR	GO; GO:0006955; P:immune response; IEA.	
DR	InterPro; IPR000973; CD4_TCAG.	
DR	InterPro; IPR007110; IG_1like.	
DR	InterPro; IPR003596; IG_V.	
DR	Pfam; PF00047; Ig; 3.	
DR	PRINTS; PR00692; CD4TCANTIGEN.	
DR	SMART; SM00406; IG; 1.	
DR	PROSITE; PS00835; IG LIKE; 2.	
SQ	SEQUENCE 455 AA; 50499 MW; AA532FD4411AA5D1 CRC64;	
Query March	34.8%; Score 1187; DB 6; Length 455;	
Blast Local Similarity	60.5%; Fred. No. 1e-82;	
Matches 251; Conservative 46; Mismatches 98; Indels 20; Gaps 5;		
QY	1 MNRGVPPHLLLVLTALLLPATQGNKVVVLGKKGDPTVLTGTASOKSIOPFMKNNSNQIK	60
Dd	1 MDPRTSLRHLPLVLQLVNLPAQTQCKVVLGKAGELAPCKASONSLSFFSKNSYTQK	60
QY	61 ILNGSGSLTITGPKSLKNDRADSRRLWDQGNFPIIKNLKLTEDSDTYICEVEDKEEVOI	120
Dd	61 ILGRHGVMHWGASNLHSRVSKESTINLMDSGFPLVIKYOLEVPDSGTVICEVDKKIEVEL	120
QY	121 LVPLGLANSPDHLLLOGOSTLTTLSEPPSSSYVCGRSRGNIQGKGKLSTLSOLELDNSG	180
Dd	121 QVFPLTVASSDPTRLILGGSLTTLTLEPGSSNPNSYVKMGKNNRKNAKSLSPVGLODNG	180

```

OY      181  TWCTYTLQNOCKYKFEKIDVLVLAFOKASSIYKKEGEQVESFPLAFVELUTSSGGL-W 233
Db      181  TWCTCTVSOAQOCTLVFNFKHLVLVAFQVBSSTVYAKKEGOMNPSFPLTFDENL--SGELSW 235
OY      240  WQERASSSKSWITFDLJKKREVSYKRVTPDPLQMGKGLPHLTLPOALPOYAGSGNLT 295
Db      239  LQKGNSSPEPSWITFDLNNKGYTVGARDDLKLRBSKALPHLTLPOALPOYAGSGNLT 296
OY      300  ALEAKTGKGLHGEVNLVYMRATOLQKNTLCEWGPSPKLMLSLKLENKAYSKREKPYW 355
Db      299  NL--TGKGYQOEYNLVYMRVTKSPNSNLCEVGLPTSPRLILSLKKNOSMVSQOQKLYT 356
OY      360  VLNPRAGMOCCLSDSGQVLTLESNFKVLPWTMSTPVPCCAPRPSKCDKTHYTELL 414
Db      357  VLPBPAGMOCCLSDGKGVLLSKYKILP---PV-----LAHANP KLL 396

```

ID	066T72	PRELIMINARY;	PRT;	354 AA.
AC	066T72;			
DT	01-JUN-2003 (TREMBLrel. 24, Created)			
DT	01-JUN-2003 (TREMBLrel. 24, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Human full-length cDNA clone CS00D1019F20 of placenta of Homo sapiens (Human) (Fragment).			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX	NCBI_TaxId=9606;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RA	Genoscope; (FEB-2003) to the EMBL/genbank/DBJ databases.			
RL	Submitted (FEB-2003) to the EMBL/genbank/DBJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	TISSUE=Placenta;			
RA	Li W.B., Gruber C., Jessee J., Polajew D.;			
RT	"Full-length cDNA libraries and normalization";			
RL	Submitted (FEB-2003) to the EMBL/genbank/DBJ databases.			
DR	EMBL; BX246827; CAD62606.1; -			
DR	GO; GO:0046821; C:extrachromosomal DNA; IEA.			
DR	InterPro; IPR007110; IG_1like.			
DR	InterPro; IPR003597; IG_c1.			
DR	InterPro; IPR003006; IG_MHC.			
DR	Pfam; PF00047; Ig_1.			
DR	SMART; SM00407; IgC1.3.			
DR	PROSITE; PSS0835; IG_LIKE.3.			
DR	PROSITE; PSS0290; IG_MHC.2.			
KW	Plasmid.			
FT	NON TER			
SQ	SEQUENCE 354 AA; 39125 MW; 238808P4D2B87A92 CRC64;			
Query Match	34.1%;	Score 1163;	DB 4;	Length 354;
Best Local Similarity	90.0%;	Pred. No. 5e-81;		
Matches 216;	Conservative 8;	Mismatches 10;	Indels 6;	Gaps 2;
0Y	392 TPVPC-APEPKSCDKTHTC-----PELLGSPVLPPEPKPDITMTISRTPEVTCVVVDV			445
Db	114 TTPPCRCRCPKSCDTPPCRCRCPAPELLGSPVLPPEPKPDITMTISRTPEVTCVVVDV			1773
0Y	446 SHEDPEVKFNMYDGVGVHNATKCRREOYNSTYRVSVLTVLHODWLNGKEYKCVSNK			505
Db	174 SHEDPEVQFKNMYDGVGVHNATKCRREOYNSTYRVSVLTVLHODWLNGKEYKCVSNK			2333
0Y	506 ALPAPIEKITSKAKGQPREPOVYTLPPSDELTKNQVSLTCLVKGFYPSDIAVMEESNGQ			565
Db	234 ALPAPIEKITSKKGQPREPOVYTLPPSDELTKNQVSLTCLVKGFYPSDIAVMEESNGQ			2939
0Y	566 PENNYKTPPVLDSDGSFFLYSKLTVDKSRWQQGNFPCSVMEHALLNHYTQKSLSPG			625
Db	294 PENNYKTPPVLDSDGSFFLYSKLTVDKSRWQQGNFPCSVMEHALLNHYTQKSLSPG			353

## RESULT 17

Q8N4Y9

ID Q8N4Y9 PRELIMINARY; PRT; 521 AA.

AC Q8N4Y9; 08N4Y9; 08N4Y9;

DT 01-OCT-2002 (TrEMBLrel. 22, Created)

DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI\_Taxid=9606;

OX NCBI\_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Primary B-Cells;

RA Strausberg R.;

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC033178; AAH33178.1; -.

DR F1R; A60764; A60764.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003597; Ig\_C1.

DR InterPro; IPR003006; Ig\_MHC.

DR InterPro; IPR003596; Ig\_v.

DR Pfam; PF00047; Ig\_4.

DR SMART; SM00407; IGL1; 3.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS00835; IG\_LIKE; 4.

DR PROSITE; PS00290; IG\_MHC; 2.

DR Hypothetical protein.

SQ SEQUENCE 521 AA; 57156 MW; 2AC7D22E7D6CAA2 CRC64;

Query Match 34.0%; Score 1160; DB 4; Length 521;

Best Local Similarity 49.2%; Pred. No. 1.5e-80;

Matches 281; Conservative 48; Mismatches 110; Indels 132; Gaps 24;

QY 117 EVGLVFLGFLANDTHLLO-GOSLTLT-----LESPPGSSPS-VQC-RSP 158  
 DB 20 EVGLV-----DSGGGVPGSSLRSLSCAASGFTVDHYEWRQAPGKPEWVGCFRKR 73  
 QY 159 RGR-----NIGGKGT-----LSVQLELDGSGTWCT-VLQNKVEFEKIDI 199  
 DB 74 AHSTTEYASVAGRTILRDSKNSVHLQMSLTKTDPTAVYCVADLEAGKDYDFPI 133  
 QY 200 ----VLAFOKASSIYKKEGEVFEFPLAFTVEKLTSGSELWQAEASSSKSWITPD 255  
 DB 134 WGRGILVTVSSAST-----KGPV-----FPLAPCSRSTSG-----GTALGCLVXD 175  
 QY 256 LKNKEYSVGRVNDPRLQWKKLPLHLTLPOALPOVAGSNTLALAEKTKL----- 308  
 DB 176 YPEPPTVS--WNSGALTSG---VH-TFPAVL-QSSGLYSLSVTVPSSSLGTOTYTC 227  
 QY 309 ---HOENILVVMRATOLQKUL-----TCEVWGPSTPKMLSLKLENKEAKVSKREKPYW 360  
 DB 228 NVNHNKSNITVDKRVLEKTPDGTHTC---PRCP-----EPR-SCDTPPPCP 271  
 QY 361 LNPBAGMOCCLSDSGOVLLESNIKVLPTWSTVPV-APBPSGCDKTHTC-----PELT 414  
 DB 272 RCPBP-----KSCDTPPCRCPEPKSCDTPPCRCAPABELL 309  
 QY 415 GGSVSLFPKPKDMLISTRPVTCVVDVSHEDBVKFMVVDGVEVNAATKPREQ 474  
 DB 310 GGSVSLFPKPKDMLISTRPVTCVVDVSHEDBVKFMVVDGVEVNAATKPREQ 369  
 QY 475 YNSTRVSVLTJLHODMLNGEKYCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSR 534  
 DB 370 FNSITFRVSVLTJLHODMLNGEKYCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSR 429  
 QY 535 DELTKQVSLTCLVKGFPSPDIADVWESNQGPNNTKTPPVLDSDGSFFLYSKLTVDKS 594  
 DB 430 EBMTKIQVSLTCLVKGFPSPDIADVWESNQGPNNTKTPPVLDSDGSFFLYSKLTVDKS 489

QY 595 RMOQGNVFCSCVMHEALNHYTKSLSPG 625  
 DB 490 RMOQGNVFCSCVMHEALNHYTKSLSPG 520

## RESULT 18

Q8RC63

ID Q8RC63 PRELIMINARY; PRT; 473 AA.

AC Q8RC63; 08RC63;

DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Hypothetical protein.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

NCBI\_Taxid=9606;

OX NCBI\_Taxid=9606;

RN [1]

RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;

RA Strausberg R.;

RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.

DR EMBL; BC025985; AAH25985.1; -.

DR GO; GO:0005507; F:copper ion binding; IEA.

DR GO; GO:0005489; F:electron transporter activity; IEA.

DR GO; GO:0006118; P:electron transport; IEA.

DR InterPro; IPR000923; BlueCu\_1.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003006; Ig\_MHC.

DR InterPro; IPR003596; Ig\_v.

DR Pfam; PF00047; Ig\_4.

DR SMART; SM00406; IGV; 1.

DR PROSITE; PS00196; COPPER BLUE; 1.

DR PROSITE; PS00835; IG\_LIKE; 4.

DR PROSITE; PS00290; IG\_MHC; 3.

DR Hypothetical protein.

SQ SEQUENCE 473 AA; 51986 MW; E29920B09BA369F5 CRC64;

Query Match 33.1%; Score 1130; DB 4; Length 473;

Best Local Similarity 51.3%; Pred. No. 2.6e-78;

Matches 260; Conservative 42; Mismatches 107; Indels 98; Gaps 15;

QY 138 SLTLTLESPGSSPVPVQ--CRSPKKNIOGKTLVS-----QLELDGSGTWCTVLQN 189  
 DB 45 SLTCTVSGSVASSTYVWGVQPPQKGLMIGTINFSNMVYSPBLRVRVMSADMSEN 104  
 QY 190 QKKVEFKIDIVLAFOKASSIYKKEGEVFEFPLAFTVEKLTSGSELWQAEASSSKSWITPD 246  
 DB 105 --SFYTKLSDVTA--DVAVYCAAGHLV-----WGRGALHGGQKLVSVSP 146  
 QY 247 -SSKSWITPDLKNKEYSVGRVNDPRLQWKKLPLHLTLPOALPOVAGSNTLALAEKTKL 302  
 DB 147 ASTKGPVFPPL---APCSRSTSESTAAAGCLVQVFPPEPVTVMNSGALTSGVHTFPVAV 202  
 QY 303 AKTGKLEHENVILVMRATOL-QKNLTCEVWGPSTPKMLSLKLENK--EAKVSKREKPYW 359  
 DB 203 LSSGSLYSLSVTVPSSSLGTIKYTCNV-----DHKPSNTKDKR----- 243  
 QY 360 VINPAGMOCCLSDSGOVLLESNIKVLPTWSTVPV-CPAPBPSGCDKTHTCPELIGSPS 418  
 DB 244 -----VESK-----YGPFGSCPA-----PEFLGSPS 265  
 QY 419 VFLFPKPKDMLISTRPVTCVVDVSHEDBVKFMVVDGVEVNAATKPREQ 478  
 DB 266 VFLFPKPKDMLISTRPVTCVVDVSHEDBVKFMVVDGVEVNAATKPREQ 325  
 QY 479 YRVSVSLTJLHODMLNGEKYCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSRDEL 538  
 DB 326 YRVSVSLTJLHODMLNGEKYCKVSNKALPAPIEKTIISAKGQPREPQVYTLPPSRDEL 385  
 QY 539 KQVSVSLTCLVKGFPSPDIADVWESNQGPNNTKTPPVLDSDGSFFLYSKLTVDKSRWQ 598  
 DB 386 KQVSVSLTCLVKGFPSPDIADVWESNQGPNNTKTPPVLDSDGSFFLYSKLTVDKSRWQ 445

Qy 599 GNVFSCSVMEALNHNHTOKSLSLSPG 625  
Db 446 GNVFSCSVMEALNHNHTOKSLSLSLG 472

## RESULT 19

P79355 PRELIMINARY; PRT; 474 AA.  
ID P79355  
AC P79355;  
DT 01-MAY-1997 (Tremblrel. 03, Created)  
DT 01-MAY-1997 (Tremblrel. 03, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE CD4 antigen precursor.  
OS Feline silverstria catus (Cat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.  
OX NCBI\_TaxID=9685;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Miyazawa T.;  
RL Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92165316; PubMed=1537604;  
RT Norimine J., Miyazawa T., Kawaguchi Y., Toyota Y., Kai C., Mikami T.;  
RT "A cDNA encoding feline CD4 has a unique repeat sequence downstream of  
RT the V-like region." (1992).  
RL Immunology 75:74-79(1992).  
DR EMBL; AB000463; BAA19124.1; -.  
DR HSSP; P01730; 1WIO.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0006955; P:immune response; IEA.  
DR InterPro; IPR000973; CD4 TCAG.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; Ig\_3.  
DR PRINTS; PRO0692; CD4TCANTIGEN.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG\_LIKE; 2.  
KM Signal.  
FT SIGNAL.  
FT CHAIN.  
SQ SEQUENCE 474 AA; 52243 MW; D946DD4BEAED00EC CRC64;

Query Match 31.8%; Score 1087; DB 6; Length 474;  
Best Local Similarity 57.6%; Pred. No. 5.3e-75;  
Matches 239; Conservative 49; Mismatches 99; Indels 28; Gaps 6;

Qy 1 MNRGVPFRLHLLVQLALPAAT-OGNKVVLGKGGDTVELTCTASQKKSIOFHKNSNOI 59  
Db 1 MNRGVPFRLHLLVQLALPAAT-OGNKVVLGKGGDTVELTCTASQKKSIOFHKNSNOI 59  
Qy 60 KILNOCG-SPL-TKGSKLNDRAADRSRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEV 118  
Db 61 KILNOCG-SPL-TKGSKLNDRAADRSRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEV 118  
Qy 119 QLVFGLTANSD-----THLQGSITLTLESPPGSSPVQCRSPRGK 161  
Db 121 ELVFGTLAKYDPSGSGSSSSSTSTSTSYVLQGSITLTLESPPGSSPVQCRSPRGK 180  
Qy 162 NIQGGKTLVSQLELDGSGTCTVYLQOKVEFKIDIVLAFKASIVYKKGGEQVEF 221  
Db 181 SKSGVHSLSTQLBQESGTCCTVQSOKTLVFNTNLVAFKRVSTVYAKGEQVEF 240  
Qy 222 SFPLAFTEKLTGSGELMWOAERASSSKWTTPDLKNEVSVKVTQDPKIQMGKPLH 281  
Db 241 SFPLAFTEKLTGSGELMWOAERASSSKWTTPDLKNEVSVKVTQDPKIQMGKPLH 281  
Qy 282 LTLPOALPOYAGSGNLTLALAKTGKGLHQBENLVVVRATQQLKULTEGVWGPTSPKMLLS 341  
Db 299 FTLNVLNVSRYAGSGNLTLVLD--KGQLOQEVKLVVVRATQQLKULTEGVWGPTSPKMLLS 356

Qy 342 LKLENKAKVS-KREKPVVNLNPEAGMOCLLSDSGVLLBSNICKLP-----TWS 391  
Db 357 LKLGQAARVS-KQOKVVRVEDAEGTQWOCLLSHDKYLLASKAVALPVLRLTWT 411

## RESULT 20

O61396 PRELIMINARY; PRT; 457 AA.  
ID O61396  
AC O61396;  
DT 01-NOV-1996 (Tremblrel. 01, Created)  
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE T-cell differentiation antigen.  
GN CD4.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=88041159; PubMed=2823269;  
RT Gorman S.D., Tourville B., Barnes J.R.;  
RT "Structure of the mouse gene encoding CD4 and an unusual transcript in  
RT brain.";  
RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).  
DR EMBL; M17080; AAA37402.1; JOINED.  
DR EMBL; M17076; AAA37402.1; JOINED.  
DR EMBL; M17077; AAA37402.1; JOINED.  
DR EMBL; M17078; AAA37402.1; JOINED.  
DR EMBL; M17079; AAA37402.1; JOINED.  
DR HSSP; P01730; 1WBR.  
DR MGD; MGI:88335; Cd4.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0006955; P:immune response; IEA.  
DR InterPro; IPR000973; CD4 TCAG.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; Ig\_2.  
DR PRINTS; PRO0692; CD4TCANTIGEN.  
DR SMART; SM00406; IGV; 1.  
DR PROSITE; PS50835; IG\_LIKE; 1.  
SQ SEQUENCE 457 AA; 51368 MW; 24AB19EDA285B5D0 CRC64;

Query Match 28.9%; Score 985.5; DB 11; Length 457;  
Best Local Similarity 50.7%; Pred. No. 3.1e-67;  
Matches 216; Conservative 66; Mismatches 121; Indels 23; Gaps 8;

Qy 1 MNRGVPFRLHLLVQLALPAAT-OGNKVVLGKGGDTVELTCTASQKKSIOFHKNSNOI 59  
Db 1 MNRGVPFRLHLLVQLALPAAT-OGNKVVLGKGGDTVELTCTASQKKSIOFHKNSNOI 59  
Qy 60 KILNOCG-SPL-TKGSKLNDRAADRSRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEV 116  
Db 61 KILNOCG-SPL-TKGSKLNDRAADRSRLMDQGNFPLIKNLKIEDSDTYICEVEDQKEV 116  
Qy 117 EVQLVFGLTANSSTHLLQGSITLTLES-PPGSSPVQCRSPRGKNIQGGKTLVSQLE 175  
Db 120 EVQLVFGLTANSSTHLLQGSITLTLES-PPGSSPVQCRSPRGKNIQGGKTLVSQLE 175  
Qy 176 LODSGTCTVYLQOKVEFKIDIVLAFKASIVYKKGGEQVEFSPPLAFTEKLTGS 235  
Db 180 VQDSDFNCTVLTQOKVFGMTLSVGFSTALTAYSGESSEFPLNFAEE--NGW 237  
Qy 236 GELMWOAERASSSKWTTPDLKNEVSVKVTQDPKIQMGKPLHLLTPOALPOYAGSG 295  
Db 238 GELMWOAERASSSKWTTPDLKNEVSVKVTQDPKIQMGKPLHLLTPOALPOYAGSG 297  
Qy 296 NLTLALAKTGKGLHQBENLVVVRATQQLKULTEGVWGPTSPKMLLSKLENKAKVSRE 355  
Db 298 NLTLALAKTGKGLHQBENLVVVRATQQLKULTEGVWGPTSPKMLLSKLENKAKVSRE 355  
Qy 356 KPVVNLNPEAGMOCLLSDSGVLLBSNICKLP-----TWS 391  
Db 357 KPVVNLNPEAGMOCLLSDSGVLLBSNICKLP-----TWS 391





```

DT 01-OCT-2003 (Tremblrel. 25, Created)
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH 11; TISSUE=Breast tumor;
RX MEDLINE=22386257; PubMed=12477932;
RA Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toshilyski S., Carinici P., Prange C.,
RA Raha S.S., Loguettano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butlerfield Y.S.,
RA Krzywnicki M.I., Skalek U., Smallus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CZECH 11; TISSUE=Breast tumor;
RA Straubeberg R.;
RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC055910; A4H5910.1; -
KM Hypothetical protein.
SQ SEQUENCE 470 AA; 51727 MW; 6D90E4DF86B090 CRC64;

Query Match 25.0%; Score 854; DB 11; Length 470;
Best Local Similarity 38.6%; Pred. No. 4.1e-57;
Matches 203; Conservative 60; Mismatches 127; Indels 136; Gaps 15;

QY 147 PGSSPSVQCSPPRGKNIQG-----GKTLVSQLELDGSGTWTCTVLQNKVYFK 196
DB 33 PGASVVKISCKA-SGYTFGYMHWVKQSHGKSLKEMIGLVNPSNGDTS---YNQK--FK 84
QY 197 IDIVLVLFQKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMWAQERASSSKRMIFDL 256
DB 85 -GKATLVVDKSSSTAYHE-----LNSLTSEDSAVYYCARYYSGSYWYFDV 129
QY 257 --KNKEVSVKRVTDQPKLQMGKLPRLTLPLQALPQYASG-----NLTLALEAK----- 304
DB 130 WGAQTYTVSSAT-----TTASVYPLVPGCGDTSGSSVTIGLGVKGFPE 175
QY 305 -----TGKLGQVNLV-----VMRATOLQKNLTCEWGPSPKMLLS 341
DB 176 PVTVMKMYGALSSGVRITVSSVLSQGFYSLSLVTPSPSTWPSQVIVCNVAHPAS----- 229
QY 342 LKLENKAKVSKREKPVWVLPNPEAGMOCCLSDSGVLLBSNIVLPTWSTPVPYCPAPR 401
DB 230 -----KTELKRIE-----RIKPSPTP 248
QY 402 KSCDKTICP--ELIGSPSYFLPPPKPDMLISRTPEVTCVVVDVSHEDPEVFKNYVD 459
DB 249 GS-----SCPRGNLTGSPVIFPPKPKDALMISLTIRKTCVVVDVSEDDPDVHVSFVD 303
QY 460 GVEVNAKTKPREQVNSYTRVSVLVTLVHODMLNGEKYCKVSNKALPAPIETKISKAK 519
DB 304 NKEVHTAMTOPREAOVNSTFRVVSALPIQHODMMRGKEFKCKVANKKALPAPIERTISKPK 363

```

```

QY 520 GQPREQVYTLPPSRDELITKNQVSLTCLVKGFPYSPDIAVMESNGQPENNYKTPPYLDS 579
DB 364 GQATPQVYTLPPSRDEKMSKKVSLTCLVNFPSSEALSVEMENNGELQGYKNTPTPLDS 423
QY 580 DGSEFFLSKLTVDKSRMQGQNVFSCSYMEALNHNHYQKLSLSPG 625
DB 424 DGYTFVLSKLTVDTSVLQGEIFPSCVVAHALNHNHYQKLSLSPG 469

RESULT 24
QY 09RIA4 PRELIMINARY; PRT; 437 AA.
ID 09RIA4
AC 09RIA4;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Gamma1 heavy chain of Mab7 (fragment).
GN IGH-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Wilde K.G., Yu X., Ekramodoullah A.K.M., Misra S.;
RT "Cloning of cDNAs encoding for anti-white pine blister rust monoclonal
RT antibody (Mab 7, its light and heavy chains) and construction of a
RT single chain antibody (scFv).";
RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF152372; AAD40243.1; -
DR PIR; B45837; B45837.
DR PDB; 1COK; 11-SEP-99.
DR PDB; 1I91; 25-DEC-02.
DR PDB; 1KCU; 11-MAY-02.
DR MDD; MGI:96446; IGH-4.
DR InterPro; IPR007110; IGH-like.
DR InterPro; IPR003006; IGH_MHC.
DR InterPro; IPR003596; IGH_v.
DR Pfam; PF00047; IGH_3.
DR SMART; SM00406; IGH_1.
DR PROSITE; PS00835; IGH_LIKE; 4.
DR PROSITE; PS00290; IGH_MHC; 1.
FT NON_TER 1
FT NON_TER 437
SQ SEQUENCE 437 AA; 48142 MW; 5C3A7B3E7D697C CRC64;

Query Match 24.6%; Score 841.5; DB 11; Length 437;
Best Local Similarity 33.2%; Pred. No. 3.4e-56;
Matches 200; Conservative 75; Mismatches 145; Indels 183; Gaps 14;

QY 30 LGKGDVVELTCTAS--QKKSIGFHKNSNQIKLNGQSFPLTKGSPKLNDRADSRSLW 87
DB 10 LVKGGSLKLSKCAAGFTFSSYMSWVRQTPREKLEWVASFSSGGLIYYDSVYGRFTIY 69
QY 88 ----DQGNFLLINKLKEDSDTYICEVEDQKEVQLLVFELTANSDTHLQGSLLTLE 144
DB 70 KDXDRNLTSLQMSLSLRBEDTAMTYC-----ARGDSAYMGPELTIVTS 112
QY 145 SPGSSPSVQCSPPRGKNIQGKTLVSQLELDGSGTWTCTVLQNKVYFKIDIVLAF 204
DB 113 AAKTTTPSYVPLAP-----GSAAGTNSMTL----- 138
QY 205 QKASSIVYKKEGQVEFSPPLAFTVEKLTGSGELMWAQERASSSKSWITFDLKNKEVSVK 264
DB 139 -----GCLVKGYPPEPYTVT-----MNSGSLSSG----- 162
QY 265 RVNQDPLQMGKSLPLHLTLPLQALPQ--VAGSGNLTLALEAKTKLHGQVNLVVMRATOL 322
DB 163 -----VH-TFPAVLQSDLYLSSVT-----VPSSTWP 189
QY 323 QKNLTCEWGPSPKMLLSLKENKAKVSKREKPVWVLPNPEAGMOCCLSDSGVLLBS 382
DB 190 SETVTCNVVAHPAS-----STKVDK-----IYPRDGCCKPCI----- 221

```

```
QY 383 NIKVLPWTSTVPCPAPBPKSCDKHTHTCPBLLGSGVFLPPPKYKDTLMISRTPEVTCV 442
DB 222 -----CTVPEVS-----SVFIFFPKPKDVLITLTPKVTCAV 253
QY 443 VDVSHDEPEVKFMYVDGVEVHNACTKPREQYNSTYRVVSVLTVLHODMLNGKEVKCY 502
DB 254 VDSKDDPEVQPSWFDVDEVHTAQTQPREQFNSTFRSVSELPIMHODMLNGKEVKCY 313
QY 503 SNKALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMES 562
DB 314 NSAAPAPAPIEKTISKTKGPRKAPQVYTIIPPKCKMAKDKVSLTCLITDFFPEDIVEMQ 373
QY 563 NGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFCGVMEALHNHYTQKSLS 622
DB 374 NGQPAENYKNTQPIMDTDSYFYYSKLVNQSNWEGAGNTFTCSVLHEDLNHTKLSH 433
QY 623 SPG 625
DB 434 SPG 436
```

## RESULT 25

```
Q8R3V9 PRELIMINARY; PRT; 469 AA.
ID Q8R3V9;
AC Q8R3V9;
DT 01-JUN-2002 (TrEMBLrel. 21, Created)
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN IGH-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC024405; AAH24405.1; -.
DR PIR; B45837; B45837.
DR MGI; MGI:96446; Igh-4.
DR InterPro; IPR007110; Igh-like.
DR InterPro; IPR003006; Igh_MHC.
DR InterPro; IPR003596; Igh_V.
DR Pfam; PF00047; Igv; 3.
DR SMART; SM00406; Igv; 1.
DR PROSITE; PS00835; IGH_LIKE; 4.
DR PROSITE; PS00290; IGH_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 469 AA; 51976 MW; 534793F15D05457 CRC64;
```

Query Match 24.6%; Score 841.5; DB 11; Length 469;

Best Local Similarity 34.6%; Pred. No. 3.8e-56;

Matches 187; Conservative 76; Mismatches 112; Indels 165; Gaps 14;

```
QY 147 PGSSPSVQC-----RSFPRK-----NIQGGKTL 170
DB 33 PGSSRLSLSCAASGFTFDYVMSWRQPPGKALEMGLFRNKANGYTTVEYASVYKGRFTIS 92
QY 171 -----VQLELQDSGTWTCVTYLNQKQVPEFIDIVLAFQKASSIVYKKKEGQY 219
DB 93 RDNSSGSLIYQNALRAEDSATYYCA--RDRSSSY-----YSGTSEFAWGGGTLV 141
QY 220 EFS-----FPLAFTVEKLTGSGELMWQAERASSSKSWITFDLKNKEVSKRVATDP 270
DB 142 TVSAAKTTPSVYPLA-----PGS-----AAQTNSMTLGLGVKGY----- 177
QY 271 KIQMGKPLHLTLPLQALPOYAGSGNLTALAEKTKLHQEV-----NLVWRATOLQKN 325
DB 178 -----FPEPVTVTWNSSGSLSSGVHTFPAVLQSDLYTLSSSVTPSSTWPSQY 224
QY 326 LTCEWNGPISPKMLSLKLENKAKVSKREKPVWVNLPEAGMQLCLSDSGQVLLSNIK 385
```

```
DB 225 VTCNVNHPAS-----STKYDK-----IVPRCCGCKPCT----- 253
QY 386 VLPTWSTVPCPAPBPKSCDKHTHTCPBLLGSGVFLPPPKYKDTLMISRTPEVTCVVDV 445
DB 254 -----CTVPEVS-----SVFIFFPKPKDVLITLTPKVTCAV 288
QY 446 SHEDPEVKFMYVDGVEVHNACTKPREQYNSTYRVVSVLTVLHODMLNGKEVKCYVSK 505
DB 289 SKDDEPEVQPSWFDVDEVHTAQTQPREQFNSTFRSVSELPIMHODMLNGKEVKCYVSK 348
QY 506 ALPAPIEKTISKAKGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQ 565
DB 349 APPAPIEKTISKTKGPRKAPQVYTIIPPKCKMAKDKVSLTCLITDFFPEDIVEMQ 408
QY 566 PENNYKTPPVLDSDGSFFLYSKLTVDKSRWQQGNVFCGVMEALHNHYTQKSLS 625
DB 409 PAENYKNTQPIMDTDSYFYYSKLVNQSNWEGAGNTFTCSVLHEDLNHTKLSHSPG 468
```

## RESULT 26

```
Q99LC4 PRELIMINARY; PRT; 463 AA.
ID Q99LC4;
AC Q99LC4;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to RIKEN CDNA 181060009 gene.
GN IGH-4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003435; AAH03435.1; -.
DR PIR; B45837; B45837.
DR HSP; P01842; 7FAB.
DR MGI; MGI:96446; Igh-4.
DR InterPro; IPR007110; Igh-like.
DR InterPro; IPR003006; Igh_MHC.
DR InterPro; IPR003596; Igh_V.
DR Pfam; PF00047; Igv; 3.
DR SMART; SM00406; Igv; 1.
DR PROSITE; PS00835; IGH_LIKE; 4.
DR PROSITE; PS00290; IGH_MHC; 1.
SQ SEQUENCE 463 AA; 51007 MW; EAA674C6BBC30783 CRC64;
```

Query Match 24.5%; Score 835; DB 11; Length 463;

Best Local Similarity 36.5%; Pred. No. 1.2e-55;

Matches 186; Conservative 73; Mismatches 123; Indels 128; Gaps 13;

```
QY 136 GQSLTLTSPSSPSVQCPSRPRK-NIQGGKTL-----VQLELQDSGTWTCVTYLN 189
DB 61 GQGLEWGEIYPSGSGTYISEKFKGKATLTTRKSSSTAYMHLSSLTSESAVYFCA----- 116
QY 190 OKKVEKIDIVLAFQKASSIVYKKKEGQVFS-----FPLAFTVEKLTGSGELMW 240
DB 117 -SSYSYSL-----FAVWGGLTVTSAAKTTPSVYPLA-----PGS----- 154
QY 241 QAERASSSKSWITFDLKNKEVSKRVATDPKIQMGKPLHLTLPLQALPOYAGSGNLT 300
DB 155 -----AAQTNSMTLGLGVKGY-----FPEPVTVTWNSSGSLSSG 188
QY 301 LEAKTKLHQEV-----NLVWRATOLQKNLTCEWNGPISPKMLSLKLENKAKVSK 355
DB 189 VHTFPAVLQSDLYTLSSSVTPSSTWPSSTVTCNVNHPAS-----STKYDK- 225
QY 356 KQWVNLNPEAGMQLCLSDSGQVLLSNIKVLPTWSTVPCPAPBPKSCDKHTCPBLLG 415
DB 236 -----IVPRCCGCKPCT-----CTVPEVS----- 254
```

```

QY 416 GSVFLFPKPKDTLMSRTPEVTCVVVDSHEDPEVKFNMYVDGVEVHNKTKPREEQ 475
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 255 --SVFIFPKPKDVLTLTPKTCVVVDISKDPEVQFSFVDDVEVHTKOTQPREEQF 312
QY 476 NSTRVVSVTLVLHODMNGEKYCKKNKLPAPIEKTIKSKAQPREPOVTLPSRD 535
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 313 NSTRVVSELPIHMODMNGEKYCKKNKLPAPIEKTIKSKAQPREPOVTLPSRD 372
QY 536 ELTNQVSLTCLVNGVPSDIAVEMESNGOPENNYKTPPLVDSGFFLYSKLTVDSR 595
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 373 QMADKXSLTCLVNGVPSDIAVEMESNGOPENNYKTPPLVDSGFFLYSKLTVDSR 432
QY 596 WQGNVSCSVMEALHNHYTKSLSPG 625
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 433 WEAGNTFCSVLHNGHHTKSLSPG 462

RESULT 27
Q99L31 PRELIMINARY; PRT; 468 AA.
ID 099L31
AC 099L31
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Similar to RIKEN cDNA 1810060009 gene.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Straube R.;
RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC003878; AA03878.1; -.
DR PDB; 2AP2; 24-NOV-99.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_3.
DR SMART; SM00406; IG_1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 468 AA; 51661 MW; 96352328B332ADB CRC64;

Query March 24.0%; Score 819; DB 11; Length 468;
Best Local Similarity 35.1%; Pred. No. 2e-54;
Matches 214; Conservative 62; Mismatches 148; Indels 186; Gaps 19;

```

```

QY 376 GQVLLSNIKVLPTWSTPVCAPAPKSCDKTHTCPPELLGSPVFLFPKPKDTLMSRT 435
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 234 -----KIEPGRPIKCP---PKC-----PAPVLLGGPSFIFPKIKDVLMSLS 277
QY 436 PEVTCVVVDVSHEDPEVKFNMYVDGVEVHNKTKPREEQNSTYRVVSVTLVHODMNG 495
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 278 PMVTCVVVDVSEDDPDQVISMFVNNVSVTLVHODMNGNSTYRVVSVTLVHODMNG 337
QY 496 KEVYCKSNKALPAPIEKTIKSKAQPREPOVTLPSRDELTNQVSLTCLVNGVPSD 555
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 338 KEVYCKSNKALPAPIEKTIKSKAQPREPOVTLPSRDELTNQVSLTCLVNGVPSD 397
QY 556 IAVEMESNGOPENNYKTPPLVDSGFFLYSKLTVDSRMOGQNVSCSVMEALHNHY 615
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 398 IYVEMTNNGKTELNTKTEPEVLDSDGFFLYSKLTVDSRMOGQNVSCSVMEALHNHY 457
QY 616 TQKSLSPG 625
    |||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 458 TTKSFSRTPG 467

RESULT 28
Q9DBL4 PRELIMINARY; PRT; 473 AA.
ID Q9DBL4
AC Q9DBL4
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE 1810060009Rik protein.
DE IGH-1 OR 1810060009RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; Pubmed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Atakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamataka I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischaer M., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schirml L.M., Staudt P., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bull C., Fletcher C., Fujita M., Gariboldi M.,
RA Guelinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombauts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohzuki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007918; BAB25349.1; -.
DR PIR; S26746; S26746.
DR HSSP; P01842; 7FAB.
DR MGD; MGI:96443; Igh-1.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_3.
DR SMART; SM00406; IG_1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
SQ SEQUENCE 473 AA; 51699 MW; 9DE057A514475FBB CRC64;

Query March 23.6%; Score 805.5; DB 11; Length 473;
Best Local Similarity 34.0%; Pred. No. 2.2e-53;

```



```

Db      90 S-RDNAKTLF-----LQMTSLSESDTAMYYCARELMLRRID-----YMG 128
Qy      137 QSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLELQDSGTWTCTVLQNKQKVEFK 196
Db      129 QCTITVSSAKTTPSVPLAPGCCDTTG----- 157
Qy      197 IDIVVLAFQKASSIYKKEGEOVESFPPLAFTVEKLTGSGELMWAQERASSSKWITFDL 256
Db      158 -----SSVTL---GCLVKGYFPESVTVT-----MNSGLSSS----- 186
Qy      257 KKEKESVVRKYQDPFLQMGKLLPLHLTLPOLPOYAGSGNLTALAEAKTGKLBQEVNLV 316
Db      187 -----VH-TPPAL-----OSGLYTWSSSVTV 207
Qy      317 MRATOLQKLTCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLPNPAQMGCLLSNSG 376
Db      208 PSSWTPSGTTCVAHPAS-----STTVDKLEP-----SG 238
Qy      377 QVLESNIKVLPTWSTPVPCEAPPEKSCDKTHTC--PELLGSPVFLFPKPKDTLMTSR 434
Db      239 PI-----STINPCF-----PCKECHKCPAPLBGGPSVFIFFPPIKQVLMISL 281
Qy      435 TPETVTCVVVDVSHEDPEVKFMYVDGVEVNAKTKPREEQYNSTYRVVSVLTVLIHOMLN 494
Db      282 TPKVTCTVVVDVSEDDPDVQISMFWNNVEVHTAQTOTREDYNSITIRVVSALPIQHODWMS 341
Qy      495 GKEYCKCKVSNKALPAPIEKTIISKAKGPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPS 554
Db      342 GKERCKCKVNNKQDLSPIERTISKIKGLVRAPOVYILPPPAQLSRKQVSLTCLVAGFNP 401
Qy      555 DIAVWESNGQPENNYKTPPVLDSDGSFFLYSKLTVDKSRMQGNVSCVMEALAHNH 614
Db      402 DISYEMWSNGHTEENYKDTAPVLDSDGSFYISKLDITKTSKMEKTDSCVNRHEGLKNY 461
Qy      615 YTKQSLSLSPG 625
Db      462 YLKKTISRSPG 472

```

## RESULT 31

```

Qy      08R3H6 PRELIMINARY; PRT; 474 AA.
AC      08R3H6;
DT      01-JUN-2002 (Tremblrel. 21, Created)
DT      01-JUN-2002 (Tremblrel. 21, Last sequence update)
DT      01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE      Hypothetical protein.
GN      AU04919.
OS      Mus musculus (Mouse).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX      NCBI_TaxID=10090;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Strauberg R.;
RL      Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC025447; AAH25447.1; -
DR      MGD; MGI:2144967; AU04919.
DR      GO; GO:0005489; P:electron transporter activity; IEA.
DR      GO; GO:0006118; P:electron transport; IEA.
DR      InterPro; IPR000345; Cytochrome BS.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003006; Ig_MHC.
DR      InterPro; IPR003596; Ig_V.
DR      Pfam; PF00047; Ig; 3.
DR      SMART; SM00406; IgV; 1.
DR      PROSITE; PS00190; CYTOCHROME_C; 1.
DR      PROSITE; PS00835; IG_LIKE; 4.
DR      PROSITE; PS00290; IG_MHC; 1.
DR      Hypothetical protein.
SQ      SEQUENCE 474 AA; 51748 MW; 8608B576CD2874A CRC64;

```

Query Match 23.3%; Score 794.5; DB 11; Length 474;

Best Local Similarity 33.4%; Pred. No. 1.6e-52; Matches 211; Conservative 72; Mismatches 150; Indels 199; Gaps 20;

```

Qy      15 QIALLPATQGNKVLQKGDPTVELTCTASQ---KKSIOFWKQ-----SNQIKILCN 64
Db      20 QVOLLQSGPE---LVKPGASVKISCRASGYAPRSK---MNMVKKRPCKGLEWIGRI 70
Qy      65 -----QSFLLTKGPKLNDRADRSLSMDGNGPFLIIKNIKIDSDTIDYIEVEDQK 115
Db      71 PRGDGDTHYSQKF--QKAKLTADKSSVTF-----LQTLTSEDSAVYFCARSD 120
Qy      116 EEVQLVFGLTANSPTHLQGSQSLTLTLESPPGSSPVQCRSPRGKNIQGGKTLVSQLE 175
Db      121 -----YG-----DYDDMGQATVTVSSAKTTPSVPLAPGCCDTTG----- 158
Qy      176 LQDSGTCTVLQNKQKVEKIDIVLAFQKASSIYKKEGEOVESFPPLAFTVEKLTGS 235
Db      159 -----SSVTL---GCLVKGYFPESVTVT----- 178
Qy      236 GELMWAERASSSKSWITFDLKNKEVSVKXVTQDPKLOMGKLLPLHLTLPOLPOYAGSG 295
Db      179 -----MNSGLSSS-----VH-TPPAL----- 195
Qy      296 NLTLAEAKTGKTHQEVNLVWRATOLQKLTCEVWGPTSPKMLSLKENKAKVSKRE 355
Db      196 -----QSGLYTWSSSVTVPSSTWSPQVTCVAHPAS-----STTVDKL 235
Qy      356 KPVWVLPNPAQMGCLLSDSGQVLESNIKVLPTWSTPVPCEAPPEKSCDKTHTC--PEL 413
Db      236 EP-----SGPI-----STINPCF-----PCKECHKCPAPNL 261
Qy      414 LGGSPVFLFPKPKDTLMTSRTPETVTCVVVDVSHEDPEVKFMYVDGVEVNAKTKPRE 473
Db      262 EGGSPVFIFFPPIKIDVLMISLTPKVTCTVVVDVSEDDPDVQISMFWNNVEVHTAQTOTHRE 321
Qy      474 QNSTYRVVSVLTVLIHOMLNKQVSLTCLVKGFYPSDIKAKGPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPS 533
Db      322 DYNSTIRVVSALPIQHODWMSGKEFKKNNKQDLSPIERTISKIKGLVRAPOVYILPP 381
Qy      534 RDELTKNQVSLTCLVKGFYPSDIKAKGPRENNYKTPPVLDSDGSFFLYSKLTVDK 593
Db      382 AEQLSRKDVSLTCLVAGFNPGRDISEWTSNGHTEENYKDTAPVLDSDGSFYISKLDIKT 441
Qy      594 SRMQGNVSCVMEALAHNHHTQKSLSLSPG 625
Db      442 SKMEKTDTSFSCVNRHEGLKNYILKKTISRSPG 473

```

## RESULT 32

```

Qy      096EYO PRELIMINARY; PRT; 613 AA.
AC      096EYO;
DT      01-DEC-2001 (Tremblrel. 19, Created)
DT      01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT      01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE      Hypothetical protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX      NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RA      Strauberg R.;
RL      Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC011857; AAH11857.1; -
DR      PIR; S15590; S15590.
DR      InterPro; IPR007110; Ig-like.
DR      InterPro; IPR003006; Ig_MHC.
DR      InterPro; IPR003596; Ig_V.
DR      Pfam; PF00047; Ig; 5.
DR      SMART; SM00406; IgV; 1.
DR      PROSITE; PS00835; IG_LIKE; 5.

```

DR PROSITE: PS00290; IG\_MHC; 3.  
 KW Hypothetical protein. 613 AA; 67273 MW; 31214203F88421E7 CRC64;  
 SQ SEQUENCE

Query Match 12.2%; Score 416.5; DB 4; Length 613;  
 Best Local Similarity 22.2%; Pred. No. 2.5e-23;  
 Matches 151; Conservative 117; Mismatches 280; Indels 131; Gaps 24;

8 RHLLVLAQLALP-----AATGKNKVLGKKDVTLETCTASQKSIQFHKNSQIKI 61  
 2 KHLMFLLLVAAFRWVLSQVQLQESGPGVLPKPSSTLTCTVSGSISYWSMIQPAQ 61  
 62 LGNQ--GSFLTKPSKLNDRADSRSL---WDGSPLLIKNLKIDSDTYICEVDQKE 116  
 62 KGLWETIGRIYTSSTINYSLSKRVMSVDTSKNPSLSKSTVTAADIVAYTC---ASQ 117  
 117 EVQLVFGFGLTANSDTHLQGSILTLTLESPPGSSPS---VQCRSPGKNIQGGKTLVS 172  
 118 PMELPTVGL-----FWGQGLTVTVSSGSASAPLPLVSC----- 153  
 173 QLELDGSGTWTCTVLONQKVERKIDIVLAFO--KASSIVYKKEGEQVEFPFLATTVK 231  
 154 ---ENSPDTSVAAGCLAODPLPDSITFSMKYKNNSDISSTRG---FPSVLRGK 203  
 232 LTSGGLMMQAEASASSKSWITFDLKNKEVSVKRVTDPKLQWCKLPLHL--TLPOLP 289  
 204 YAATSOVLLPSKV-----MGTDEHVCKVQHFNKKEKNVPLVIAELPPKVS 253  
 290 QVA-----GSGNLTLALEAKTKLQEVNLVVMR-----ATOLQKNTCEVWGPT 334  
 254 VFVPRDFGFGNPKSKLIGQATGFSRQIVSWLRGQVSGVTTDQVAEAKESGPT 313  
 335 SPFLMISLKLKNEKAVSKREKVVYLNPAQMWOCCLSDSGQVLE--SNIKVLPWTSP 393  
 314 TYKVTSTLTIKESD-----WL---SQSMFTCRVDRGLTFQONASMCVPPQDPA 360  
 394 VPPRAPERSCDHTHCPELLGGSVFLPPKPKDITMISRTPEVTCVAVDVSHEDPEVK 453  
 361 I-----RVFALPPS--FASIFLTGSTKLTLVTDLTYY--SVT 395  
 454 FNNVVDGVEVHNAKTPREQVNSTYRVSVLTVLHODWLNKGYCKVSNKALPAPIRK 513  
 396 ISVTRNGEAVKTHNTISESHPNATPSAVGEASICEEDMNSGERFTCTVHTDLPSPHQ 455  
 514 TISKAGQRP-REQVYTLTPSRDELTKNOVSLTCLVKGFPSPDIAYEWSNQRP--ENN 569  
 456 TISRPGVALHRDPVYLLPPAREQLMRESATITCLVTGFSPADVFVQWQROGQPLSPK 515  
 570 YKTPPVLQD--SDGSFPLYSKLTVDKSRMOQGNVFSQVMEALHNHYOKSLSPGLQ 627  
 516 YVTSAPWPEQAGRFYAHSLTVSEBEMNTGETYTCVAHEALPNKVTERTYDKS---- 571  
 628 LDETCAEQDDELGLMTT 646  
 572 -TEGEVSADDEGFENLMAT 589

RESULT 33  
 Q8WUK1 PRELIMINARY; PRT; 613 AA.  
 AC 08WUK1;  
 DT 01-MAR-2002 (Tremblrel. 20, Created)  
 DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)  
 DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.  
 OC NCBI\_Taxid=9606;  
 RN 1;  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=tonsil;  
 DT Strauberg R.;

RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC020240; AAH02040.1; --  
 DR PIR; P01020; P01020.  
 DR PIR; S15590; S15590.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG\_5.  
 DR SMART; SM00406; IG\_V.1.  
 DR PROSITE; PS00835; IG\_LIKE; 5.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 613 AA; 67296 MW; 60C7F5950671B315 CRC64;

Query Match 11.9%; Score 406.5; DB 4; Length 613;  
 Best Local Similarity 22.6%; Pred. No. 1.5e-22;  
 Matches 150; Conservative 119; Mismatches 252; Indels 143; Gaps 27;

25 GNKVVLGKKGDVETLTCTAS--QKSIQFHW-----KSNQIKILGNQS-----FLTK 71  
 27 GGGV--QPRGLRSLSCAASGFTSSYGHWVRQAPGKLENAVAVSYDGSNKYIADSVK 84  
 72 GSKLNDRADSRSLMDQGNFPLIKNLKIDSDTYICEVDQKEVQLVFGLTANSPT 131  
 85 GFTIS--RDNSTNTLYLQWN-----SLRAEDTAVYVC--AKDMSEGV-----ET 125  
 132 HLLQGSILTLTLESPPGSSPS---VQCRSPGKNIQGGKTLVSQLELDGSGTWTCTVL 187  
 126 FDIWQGMWTVYSSGSASAPLPLVSC-----ENSPDTSVA 164  
 188 QNOKVEFKIDIVLAFO--KASSIVYKKEGEQVEFPFLATVEKLTSGSGLMMQAEAS 246  
 165 VGLADQFLPDSITTSMTKYNNSDISSTRG---FPSVLGCKTAATSOVLLPSKV- 217  
 247 SSKSWITFDLKNKEVSVKRVTDPKLQWCKLPLHL--TLPOLPYA-----GSGNL 297  
 218 ---MGTDEHVCKVQHFNKKEKNVPLVIAELPKVSVFVPRDFGFGNPKR 268  
 298 TLALEAKTKLQEVNLVVMR-----ATOLQKNTCEVWGPTSPKMLSLKLENKA 349  
 269 SKLIGQATGFSRQIVSWLRGQVSGVTTDQVAEAKESGPTTYKVTSTLTIKESD- 327  
 350 KVSKEKPVVYLNPAQMWOCCLSDSGQVLE--SNIKVLPWTSPVPCPAPERKSCDXT 408  
 328 ---WL---SQSMFTCRVDRGLTFQONASMCVPPQDPA----- 361  
 409 TCEPLLGGSVFLPPKPKDITMISRTPEVTCVAVDVSHEDPEVKFNNVVDGVEVHNAKT 468  
 362 ---RVFALPPS--FASIFLTGSTKLTLVTDLTYY--SVTISVTRNGEAVKTHT 410  
 469 KREEQVNSTYRVSVLTVLHODWLNKGYCKVSNKALPAPIRKISAKQRP-REPV 527  
 411 NISESHPNATPSAVGEASICEEDMNSGERFTCTVHTDLPSPKQITSPKGVALHRPV 470  
 528 YTLPSRDELTKNOVSLTCLVKGFPSPDIAYEWSNQRP--ENNYKTPPVLQD--SDGS 582  
 471 YLLPPAREQLMRESATITCLVTGFSPADVFVQWQROGQPLSPKVTYAPWPEQAPR 530  
 583 FFLYSKLTVDKSRMOQGNVFSQVMEALHNHYOKSLSPGLQDETCAEQDDELQ 642  
 531 YFAHSITLVSEBEMNTGETYTCVAHEALPNKVTERTYDKS-----TEGEVSADDEGFEN 585  
 643 LMTT 646  
 586 LMAT 589

RESULT 34  
 077596 PRELIMINARY; PRT; 86 AA.  
 ID 077596;  
 AC 077596;  
 DT 01-NOV-1998 (Tremblrel. 08, Created)  
 DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)

DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)  
OS (Fragment).  
OS Mandrillae sphinx (Mandrill) (Papio sphinx).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Mandrillae.  
OX NCBI\_TaxID=9561;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98320644; PubMed=9656488;  
RA Harris E.E., Disotell T.R.;  
RT "Nuclear gene trees and the phylogenetic relationships of the  
mangabeys (primates: Papionini).";  
RL Mol. Biol. Evol. 15:892-900(1998).  
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL  
RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY  
SIMILARITY).  
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
ONE V-LIKE AND 1 C2-LIKE DOMAIN.  
DR EMBL; AF057386; AAC25130.1; -.  
DR HSSP; P01730; 1CDY.  
DR GO; GO:0016021; C:Integral to membrane; IEA.  
DR InterPro; IPR007110; IG-like.  
KM Immunoglobulin domain; T-cell; MHC; Transmembrane.  
FT NON\_TER 1  
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.  
FT DISULFID 20 >86 IG-LIKE C2-TYPE DOMAIN.  
FT NON\_TER 49 78 BY SIMILARITY.  
FT NON\_TER 86 86  
SQ SEQUENCE 86 AA; 9406 MW; 2BD97A9EB19582AB CRC64;  
Query Match 11.7%; Score 400; DB 6; Length 86;  
Best Local Similarity 89.4%; Pred. No. 2.7e-23;  
Matches 76; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
QY 107 YICVEDEKKEVQLVGLTANSPTHLQGSSTLTLESPPGSSPSVQCRPRGKNIQGG 166  
DB 1 YICVEDEKKEVQLVGLTANSPTHLQGSSTLTLESPPGSSPSVQCRPRGKNIQGG 60  
QY 167 KTLVSQLELDQSGTWTCTVLOK 191  
DB 61 RTLSVPLEQRDQSGTWTCTVSQDK 85  
RESULT 35  
ID 077597 PRELIMINARY; PRT; 86 AA.  
AC 077597;  
DT 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)  
OS (Fragment).  
OS Mandrillae leucophaeus (Drill) (Papio leucophaeus).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Mandrillae.  
OX NCBI\_TaxID=9568;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=98320644; PubMed=9656488;  
RA Harris E.E., Disotell T.R.;  
RT "Nuclear gene trees and the phylogenetic relationships of the  
mangabeys (primates: Papionini).";  
RL Mol. Biol. Evol. 15:892-900(1998).  
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL  
RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY  
SIMILARITY).  
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS  
ONE V-LIKE AND 1 C2-LIKE DOMAIN.  
DR EMBL; AF057387; AAC25131.1; -.  
DR HSSP; P01730; 1CDY.  
DR GO; GO:0016021; C:Integral to membrane; IEA.  
DR InterPro; IPR007110; IG-like.  
KM Immunoglobulin domain; T-cell; MHC; Transmembrane.  
FT NON\_TER 1  
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.  
FT DISULFID 20 >86 IG-LIKE C2-TYPE DOMAIN.  
FT NON\_TER 49 78 BY SIMILARITY.  
FT NON\_TER 86 86  
SQ SEQUENCE 86 AA; 9406 MW; 2BD97A9EB19582AB CRC64;  
Query Match 11.7%; Score 400; DB 6; Length 86;  
Best Local Similarity 89.4%; Pred. No. 2.7e-23;  
Matches 76; Conservative 6; Mismatches 3; Indels 0; Gaps 0;  
QY 107 YICVEDEKKEVQLVGLTANSPTHLQGSSTLTLESPPGSSPSVQCRPRGKNIQGG 166  
DB 1 YICVEDEKKEVQLVGLTANSPTHLQGSSTLTLESPPGSSPSVQCRPRGKNIQGG 60  
QY 167 KTLVSQLELDQSGTWTCTVLOK 191  
DB 61 RTLSVPLEQRDQSGTWTCTVSQDK 85  
RESULT 36  
ID 077MT6 PRELIMINARY; PRT; 614 AA.  
AC 077MT6;  
DT 01-OCT-2003 (Tremblrel. 25, Created)  
DT 01-OCT-2003 (Tremblrel. 25, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;  
RX MEDLINE=22388257; PubMed=12477937;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
RA Brownstein M.J., Ueda T.B., Tomihata S., Carninci P., Prange C.,  
RA Rana S.S., Lochuano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
RA Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahey J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
RA Krzywinski M.I., Skalske U., Smallus D.E., Schnerch A., Schein J.E.,  
RA Jones S.J., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length human  
and mouse cDNA sequences";  
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX STRAIN=C57BL/6NCR; TISSUE=Hematopoietic Stem Cell;  
RA Strausberg R.;  
RL Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC053409; AAH53409.1; -.  
KM Hypothetical protein.  
SQ SEQUENCE 614 AA; 67746 MW; 839BAF3B8D124F89 CRC64;  
Query Match 11.7%; Score 400; DB 11; Length 614;

Best Local Similarity 22.5%; Pred. No. 4,6e-22;  
Matches 156; Conservative 98; Mismatches 232; Indels 206; Gaps 28;

```

QY 30 LGKKGDTVELTCTAOKKSIQFMKN-----SNQIKILGN-----QGSFLTKGP 73
DB 30 LVPKASVAKISCKAS-GYAFSSSMNMWVKORPGKIGIEWGRVYPGDGTNYNGKF--KKG 86
QY 74 SKLN-DRAASRSLMUGNPLIILKLIKEDSTTYICEVEDQKEEYQVLVFGLTANSDBH 132
DB 87 ATLTDKSSSTAYMQ-----LSSLTSEDSAVYFC-ARDGSSYRFAYWG----- 129
QY 133 LLOGGSLTLTLLESP--PGSSPSVQCRSP-RGNVIOGQKTLVSQQLBODSGTWTCTVLON 189
DB 130 --OGTIVTASBSQSPFNVPFLVSCSPSLSDKULVAMGLARPLSTISFTWN----- 181
QY 190 QKKVEFKIDIVLAFQKASSIYVKKGEQVEFSPLAFVTEKLTGSGELMWQAEARASSK 249
DB 182 -----YQNNAEVIQ-----GIR 193
QY 250 SWITPDLKNKESVKNVTDPRKIQM-----GKKLPLHLTPQALPOYAGSG 295
DB 194 TPTLTGTGKYLTATSOVLSPKSLLEGSEYLVCKIHGKKNDLHPVP----- 243
QY 296 NLTLALEAKTGKLGHOENLVV-----MRATOLQKNLTCEVMGPTSPKMLSLKENKA 349
DB 244 -----AVAEKNPVNVVPPRDRGSGAPRKSKLICENTFTKPTIVSWLXKQKLV 295
QY 350 KVSREKRPVWVILN---PEA-----GMQCLISDSGOVLLESNIKVL 388
DB 296 ESGFTTDPVTIENKSGSTPQYVKYISTLTISEIDMLNLYTCAVDRHGLFLK----- 348
QY 389 TWSTPVPKAPREKSCDKHTCELLGGS---VLFPRKPKDITLMISTPVTVCVVD 444
DB 349 -----NVSSSTC---AASPTDILTFTIPSPAD-TFLSKSANLTLCLVSN 388
QY 445 VSHEDPEVKMNVYDGEVNAKTKPREEQNSTYRVSVLYLTHODMLNGEKYKCVSN 504
DB 389 LATYE-TLINSWASQSGEPLFTKIKIMESHNGTFKAGVASCVEDMNRKKEFVCTVTH 447
QY 505 KALPARIKTIKAKQPRE---POVYTLPPSRDEL-T-KNOVSLTCLVKGFPDIAV 558
DB 448 RDLPSQKPKFISK---PNEVHGKPPAVYLLPPAREQNLRESATYCLVKGSPADISV 503
QY 559 EWSNQC--PENNYKTPPLD--SDGSFELYSKLTVDKSRMGGQNVFSGSVNHEALHN 614
DB 504 QWLQRQQLPQEKYVTSAPMEPGARPGFTHTSLTVEEMNSGETTYTCVSHALPHL 563
QY 615 YTKSLSLSPGLQDTECAEADGELDLMTT 646
DB 564 VTERIVDKS-----TGEVNAEBEGPENLMTT 590

RESULT 37
QY 077594 PRELIMINARY; PRT; 86 AA.
DB 077594;
AC 077594;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Cercopithecus mitis.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OC NCBI_TaxID=36225;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).

```

```

CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL; AF057381; AAC25125.1; -.
DR HSSP; P01730; 1CDY.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR007110; I9-like.
KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT DOMAIN 1 1
FT 1 19 IG-LIKE V-TYPE DOMAIN.
FT 20 86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86
SQ SEQUENCE 86 AA; 9417 MW; 371CA39EF58182AB CRC64;

Query Match 11.6%; Score 397; DB 6; Length 86;
Best Local Similarity 89.4%; Pred. No. 4,6e-23;
Matches 76; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEDQKEEYQVLVFGLTANSDBHLLQGSLTLTLLESPGSSPSVQCRSPRGKNIQCG 166
DB 1 YICEVEDKKEEYELLVFGLTANSDBHLLQGSLTLTLLESPGSSPSVQCRSPRGKNIQCG 60
QY 167 KTLVSQLELDQSGTWTCTVLDNOK 191
DB 61 RLISVQLERODSGTWTCTVSQDOK 85

RESULT 38
QY 077599 PRELIMINARY; PRT; 86 AA.
DB 077599;
AC 077599;
DT 01-NOV-1998 (Tremblrel. 08, Created)
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Cercopithecus gelada (Gelada baboon).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Theropithecus.
OC NCBI_TaxID=9565;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
DE -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
DE RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
DE SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
DR EMBL; AF057389; AAC25133.1; -.
DR HSSP; P01730; 1CDY.
DR GO; GO:0016021; C:integral to membrane; IEA.
DR InterPro; IPR007110; I9-like.
KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
FT DOMAIN 1 1
FT 1 19 IG-LIKE V-TYPE DOMAIN.
FT 20 86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 BY SIMILARITY.
FT NON TER 86
SQ SEQUENCE 86 AA; 9420 MW; 2BD97A98574EE9AB CRC64;

Query Match 11.6%; Score 397; DB 6; Length 86;

```



Best Local Similarity 88.2%; Pred. No. 4.6e-23;  
Matches 75; Conservative 7; Mismatches 3; Indels 0; Gaps 0;

QY 107 YICVEVDQKEVQLLVFGLTNSDTHLQGSLLTLESPPGSSPVQCSPPRKNIQGG 166  
DB 1 YICVEVDKEKEVELLVFGLTNSDTHLQGSLLTLESPPGSSPVQCSPPRKNIQGG 60

QY 167 KTLVSQLELODSDGTWCTVLQONK 191  
DB 61 KTLVSQLELODSDGTWCTVLQONK 85

## RESULT 39

ID 077595 PRELIMINARY; PRT; 86 AA.  
AC 077595;  
DT 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).  
OS Cercopithecus galicetus chrysogaester.  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Cercopithecus.  
OX NCBI\_TaxID=75569;

RP SEQUENCE FROM N.A.  
RX MEDLINE=98320644; PubMed=9656488;  
RA Harris E.E., Disotell T.R.;  
RT "Nuclear gene trees and the phylogenetic relationships of the mangabeys (primates: Papionini).";  
RL Mol. Biol. Evol. 15:892-900(1998).  
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY SIMILARITY).  
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS ONE V-LIKE AND 1 C2-LIKE DOMAIN.

DR EMBL; AF057382; AAC25126.1; -.  
DR HSSP; P01730; 1CDY.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR InterPro; IPR007110; Ig-like.  
KM Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.  
FT NON\_TER 1  
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.  
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DISULFID 49 78 BY SIMILARITY.  
FT NON\_TER 86 86  
SQ SEQUENCE 86 AA; 9419 MW; A9D97A98E19582BE CRC64;

Query Match 11.6%; Score 395; DB 6; Length 86;  
Best Local Similarity 88.2%; Pred. No. 6.6e-23;  
Matches 75; Conservative 6; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICVEVDQKEVQLLVFGLTNSDTHLQGSLLTLESPPGSSPVQCSPPRKNIQGG 166  
DB 1 YICVEVDKEKEVELLVFGLTNSDTHLQGSLLTLESPPGSSPVQCSPPRKNIQGG 60  
QY 167 KTLVSQLELODSDGTWCTVLQONK 191  
DB 61 KTLVSQLELODSDGTWCTVLQONK 85

## RESULT 40

ID 077598 PRELIMINARY; PRT; 86 AA.  
AC 077598;  
DT 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)

DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).  
OS Papio sp. (baboon).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
OC Cercopithecinae; Papio.  
OX NCBI\_TaxID=61183;

RP SEQUENCE FROM N.A.  
RX MEDLINE=98320644; PubMed=9656488;  
RA Harris E.E., Disotell T.R.;  
RT "Nuclear gene trees and the phylogenetic relationships of the mangabeys (primates: Papionini).";  
RL Mol. Biol. Evol. 15:892-900(1998).  
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY SIMILARITY).  
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS ONE V-LIKE AND 1 C2-LIKE DOMAIN.

DR EMBL; AF057388; AAC25132.1; -.  
DR HSSP; P01730; 1CDY.  
DR GO; GO:0016021; C:integral to membrane; IEA.  
DR InterPro; IPR007110; Ig-like.  
KM Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.  
FT NON\_TER 1  
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.  
FT DOMAIN 20 >86 IG-LIKE C2-TYPE DOMAIN.  
FT CARBOHYD 79 79 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT DISULFID 49 78 BY SIMILARITY.  
FT NON\_TER 86 86  
SQ SEQUENCE 86 AA; 9433 MW; A9D97A98574EE9BE CRC64;

Query Match 11.5%; Score 392; DB 6; Length 86;  
Best Local Similarity 87.1%; Pred. No. 1.1e-22;  
Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICVEVDQKEVQLLVFGLTNSDTHLQGSLLTLESPPGSSPVQCSPPRKNIQGG 166  
DB 1 YICVEVDKEKEVELLVFGLTNSDTHLQGSLLTLESPPGSSPVQCSPPRKNIQGG 60  
QY 167 KTLVSQLELODSDGTWCTVLQONK 191  
DB 61 KTLVSQLELODSDGTWCTVLQONK 85

QY 107 YICVEVDQKEVQLLVFGLTNSDTHLQGSLLTLESPPGSSPVQCSPPRKNIQGG 166  
DB 1 YICVEVDKEKEVELLVFGLTNSDTHLQGSLLTLESPPGSSPVQCSPPRKNIQGG 60  
QY 167 KTLVSQLELODSDGTWCTVLQONK 191  
DB 61 KTLVSQLELODSDGTWCTVLQONK 85

## RESULT 41

ID 096BB9 PRELIMINARY; PRT; 597 AA.  
AC 096BB9;  
DT 01-DEC-2001 (Tremblrel. 19, Created)  
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)  
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.  
OX NCBI\_TaxID=9606;  
[1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=B-cell;  
RA Strausberg R.;  
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC015760; AAH15760.1; -.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003006; Ig\_MHC.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig\_5.  
DR SMART; SM00406; IG\_1.  
DR PROSITE; PSS0835; IG\_LIKE; 5.  
DR PROSITE; PSS00290; IG\_MHC; 3.  
KM Hypothetical protein.

Seq	SEQUENCE	597 AA;	65039 MW;	4FC63AD8BEECE263D9	CRC64;
Qy	Query Match	11.5%;	Score 391.5;	DB 4;	Length 597;
Db	Best Local Similarity	22.4%;	Pred. No. 26-22;		
	Matches 143;	Conservative 116;	Mis.matches 248;	Indels 131;	Gaps 24
Qy	30 LGKKGDVTELCTAS--QKKSIOFPMKNSNQK-----ILGNQSF-----TKGPKLT	76			
Db	30 LVQPGSRRLSCASGFSFSSYAMWROAPGKGLBWSAISGSGSTYYADSVKGRFTI	89			
Qy	77 NDRADRSRLMDQGNFPLIKNLKLTEDSDPTICEVEDQKEEVOLVFGLTANSDFHLLQG	136			
Db	90 S-RDMSRRTLYQNM-----SLRADETAVVYCAADPRG-----YASAGVYTRDDYWG	135			
Qy	137 QSLTTLTSPGSSPS---VQCRSPRKNIOGKGLTVSGLDLQSDGTWCTVYLONQK	192			
Db	136 QGLTVTVSSGSAFAETLPFLVSC-----EKSPEDTSSVAAGCLA	174			
Qy	193 VEFKLDIYVLAFO--KASSTIVYKKEGEQVSEFPLATFVEKLTGSGSGLMWAQEAASSKSM	251			
Db	175 QDFLPDSITFSKWKYKNSDISSTRG-----FPSVLRGKVAATSOVLLPSKDV-----	222			
Qy	252 ITFDLKNKEVSKRYTODPKLOMGKKLPLHL--TLFQALPYQA-----GSGNLTALAE	302			
Db	223 ----WQGDENHVCVQHQHNGKKEKNVLPVLAELRPKXSVFPRPDGFFGNPRKSKLIC	278			
Qy	303 AKTGKLDHGEVNLVNR-----ATOLQKNLTCEVWGFTSPKMLSLKLENKEAVSKR	354			
Db	279 QATGSPRQIQVSMRLREGKQVGSVTTDQVQAEKSGSPTYKVTSTLTKESD-----	332			
Qy	355 EKPVVAVLPEAGMOCILSDSQVLL--SNIKVLPFTWSFPVPCPAEPRKSCDKTHCPRL	413			
Db	333 ----WL-----SOGMFCRVDHRLGLTQQNASMSVCVPDDTAI-----	366			
Qy	414 LGSPSEVLPFPKPKOTLMTSTPEVTCVAVDVDSHEDPEYKFMVYVDGVEHNAKTRPEE	473			
Db	367 ----RVFALPFS--PASIFLTRKTKLCLVTLDTLTYD--STISWTRONGEAVKHTHTMSES	420			
Qy	474 QYNSTYRVAVSVLTVLHQDMLNGEKYKCVSNKALPAPIKTSISKAGQP--REFOVYTLPP	532			
Db	421 HPNATFSAVGEASICEDDWNSGERFCTVTHDLPPLKQTIISRPGVALHPRDVLPLP	480			
Qy	533 SRDELTKNOVSLTCLVKCFYSDIAVEMESNGQP--ENNYKTPPVLD--SDGSEFVLS	587			
Db	481 ARBQNLNRESAITTCLVTFGSPADVPEVQMMQRQPLSPKRYTSAEMPEQAGRYFAHS	540			
Qy	588 KLTVDKSRMQOGNPFSCSYMEHALNHHYQKSLSPG 625				
Db	541 ILTVSEEMWTGETYTCVVAHEALPRLVTERIVDKSTG 578				
RESULT 42					
077601	PRELIMINARY;	PRT;	86 AA.		
AC	077601				
DT	01-NOV-1998 (TREMBLrel. 08, Created)				
DT	01-NOV-1998 (TREMBLrel. 08, last sequence update)				
DT	01-OCT-2003 (TREMBLrel. 25, last annotation update)				
DE	T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3)				
OS	(Fragment).				
OS	Lophocobus albigena albigena.				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;				
OC	Cercopitheciinae; Lophocobus.				
OX	NCBI_TaxID=75568;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=98320644; PubMed=9656489;				
RA	Harris E.E., Disocell T.R.;				
RT	"Nuclear gene trees and the phylogenetic relationships of the				
RT	mangabeys (primates: Papionini)";				
Mol. Biol. Evol.	15:892-900(1998).				
CC	-!- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL				

```

CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH p56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.
CC EMBL: AF057391; AAC25135.1; -.
CC HSSP: P01730; ICDF.
CC DR GO: GO:0016021; C: integral to membrane; IEA.
CC DR InterPro: IPR007110; Ig-like.
CC KW Immunoglobulin domain; T-cell; MHC; Transmembrane.
CC FT NON TER 1 1
CC FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
CC FT DISUPRD 20 >86 IG-LIKE C2-TYPE DOMAIN.
CC FT NON TER 49 78 BY SIMILARITY.
CC FT NON TER 86 86
CC SQ SEQUENCE 86 AA; 9463 MW; 2BD97A88464FE9AB CRC64;

Query Match 11.4%; Score 388; DB 6; Length 86;
Best Local Similarity 87.1%; Pred. No. 2.3e-22;
Matches 74; Conservative 7; Mismatches 4; Indels 0; Gaps 0;

QY 107 YICEVEDEKKEVQLVGLTRANSTHLLGQSLLTLESPPGSSPSVQCSPPKKNIQGG 166
DB 1 YICEVEDEKKEVQLVGLTRANSTHLLGQSLLTLESPPGSSPSVQCSPPKKNIQGG 60
QY 167 KTLVSQLELDGSGTWCTCTVLQXK 191
DB 61 RILSVPOLERODSGTWCTCTVLSQDOK 85

RESULT 43
096GA6 PRELIMINARY; PRT; 614 AA.
AC 096GA6:
DT 01-DEC-2001 (TRENBLrel. 19, Created)
DT 01-DEC-2001 (TRENBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=B-cell;
RA Strausberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; BC009851; AAH09851.1; -.
DR PIR; S15590; S15590.
DR GO; GO:0005622; C: intracellular; IEA.
DR GO; GO:0003700; F: transcription factor activity; IEA.
DR GO; GO:0006355; P: regulation of transcription, DNA-dependent; IEA.
DR InterPro: IPR000005; HTHRAC.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003006; Ig-MHC.
DR InterPro: IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 5.
DR SMART; SM00406; IGV. 1.
DR PROSITE; PS00041; HTH_ARAC_FAMILY_1; 1.
DR PROSITE; PS00835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein..
SQ SEQUENCE 614 AA; 67921 MW; 55EF536E77AA9BBB CRC64;

Query Match 11.4%; Score 388; DB 4; Length 614;
Best Local Similarity 22.0%; Pred. No. 3.8e-21;
Matches 150; Conservative 116; Mismatches 272; Indels 144; Gaps 26;

QY 11 LVVLQALALPATQGNKVLG-----KKGDYELTCTASQKSI--QFHKNSNQIKIL-- 62
DB 7 ILFLVAALTDAYSQMLVQSGAEVKKGTSSVYKVCCKASGYTFYRKLHWQAQPAQALEW 66

```

```

QY 63 -----GNQGSFLTKGPSKLTNDRADSRSLMDQGNPFLITKLIKIEDSDTYICEVEDQ 114
DB 67 MGWITPFGN-----TNYAQKFORVITTRDR-SMNTAYMELSLRSRSDTAMTYC----- 115
QY 115 KEVQLVFGLTAN-SDTHLQGSLLTLESPPGSSP-----VQCRSPRKNIQGGKTL 169
DB 116 -----ARGSSSWDDAFDIMGQITWTVSSGSASAPLTPLVSC----- 154
QY 170 SVSGLQELDQSGMTCTVQNGKQVEFKIDIVLAFQ-KASSIVKKEGEVFPPLAFT 228
DB 155 -----ENSPSDTSSVAVGCLAQDFLPDSITFSKYNKNSDISTRG-----PPSVLR 201
QY 229 VEKLTGSGELMWQAEKSSSSKSWITFDLKNKEVSKRYTQDPKLOMGKPLHL--TLPO 286
DB 202 GGTAAATQVLLPBDV-----WGTDENHYCKVQHPNKNKEGNVFLPIALP 251
QY 287 ALPOYA-----GSGNLTALAEKTKLHGVNLVNR-----ATOLQKNTCEVW 331
DB 252 KVSFVFPDRDGFNGPRKSKLICATGFSPRQIQVSWLREGKQVSGVTTDQVQAEAKES 311
QY 332 GPTSPKMLSLKLENKEAKSKREKPVVNLPEAGMQLSDSGQVILE-SNIKVLPTW 390
DB 312 GPTTYKVTSTLTIESD-----WL---SQSMFTCRVDHRLGLTFQGNASSMCVDPQ 358
QY 391 STVPCEPAPBPKSCDKHTTCPELLGSPVFLFPKPKDITMISRTPEVTCVVDVSHEDP 450
DB 359 DTAI-----RVFAIPPS-FASITLTSTKTLCTVLTITTYD- 393
QY 451 EVKENWYVDGVEVHNATKPREEOYNSTYRVSVLTVLHODMLNGKEKCVSKALPAP 510
DB 394 SVTISMTQNGEAVKTNHINSESHPNATFSAVGEISICEDDMNGSERFTCTVHTDLPSP 453
QY 511 IEKITSKAKGP-REPOVYTLPPSRDEL-T-KQVSLTCLVGFPSDIAVWESNGQP-- 566
DB 454 LKQITSRKGVALHRPDVLLPPAREQJLINESATITCLVGFSPADVFQMGORGPLS 513
QY 567 ENNKKTTPPVLD--SDGSFPLYSKLTVDKSRMOQGNVSCVMSEALHNHTOKSLSP 624
DB 514 PEKVTASAPMEPQAPRGVFAHSILITVSEEMNGERTYCVVAHEALPNRTERTVDKS- 572
QY 625 GLQUDETCAEADGELDGLMTT 646
DB 573 -----TEGEVSADEGEFENLMAT 590

RESULT 44
QY 077600 PRELIMINARY; PRT; 86 AA.
AC 077600;
DT 01-NOV-1998 (Tremblrel. 08. Created)
DT 01-NOV-1998 (Tremblrel. 08. Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25. Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/LEU-3) (Fragment).
OS Lophocebus aetherius.
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Lophocebus.
NCBI_TaxID=75566;
OX
RN
RP SEQUENCE FROM N.A.
RX MEDLINE=98320644; PubMed=9556488;
RA Harris E.E., Disocell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: ACCESSORY PROTEIN FOR MHC CLASS-II ANTIGEN/T-CELL
CC RECEPTOR INTERACTION. MAY REGULATE T-CELL ACTIVATION (BY
CC SIMILARITY).
CC -1- SUBUNIT: ASSOCIATES WITH P56-LCK (BY SIMILARITY).
CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE IMMUNOGLOBULIN SUPERFAMILY. CONTAINS
CC ONE V-LIKE AND 1 C2-LIKE DOMAIN.

```

```

DR EMBL; AF057390; AAC25134.1; -.
DR HSSP; P01730; 1CDY.
DR GO; GO:0016021; C: integral to membrane; IEA.
DR InterPro; IPR007110; Ig-like.
DR Immunoglobulin domain; Glycoprotein; T-cell; MHC; Transmembrane.
FT NON_TER 1
FT DOMAIN <1 19 IG-LIKE V-TYPE DOMAIN.
FT CARBOHYD 20 >86 IG-LIKE C2-TYPE DOMAIN.
FT DISULFID 49 78 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT TER 86 BY SIMILARITY.
SQ SEQUENCE 86 AA; 9476 MW; A9D97A88464FE93BE CRC64;

Query Match 11.2%; Score 383; DB 6; Length 86;
Best Local Similarity 85.9%; Pred. No. 5.5e-22;
Matches 73; Conservative 7; Mismatches 5; Indels 0; Gaps 0;

QY 107 YICEVEDQKEVQLVFGLTANSDTHLQGSLLTLESPPGSSPVSQCRSPRKNIQGG 166
DB 1 YICEVEDQKEVQLVFGLTANSDTHLQGSLLTLESPPGSSPVSQCRSPRKNIQGG 60
QY 167 KTLVSQLELDQSGMTCTVQNGK 191
DB 61 RTLSVPQLERQDSGWTCTVNSQDOK 85

RESULT 45
QY 096A06 PRELIMINARY; PRT; 618 AA.
AC 096A06;
DT 01-DEC-2001 (Tremblrel. 19. Created)
DT 01-DEC-2001 (Tremblrel. 19. Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25. Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
OX
RN
RP SEQUENCE FROM N.A.
RX
RC TISSUE=Lymph;
RA Strauberg R.;
RL Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC017356; AARI17356.1; -.
DR PIR; S15590; S15590.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; Ig_5.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 5.
DR PROSITE; PS00290; IG_MHC; 3.
KW Hypothetical protein.
SQ SEQUENCE 618 AA; 67758 MW; 96DBD4C7C96E0A6 CRC64;

Query Match 11.2%; Score 382; DB 4; Length 618;
Best Local Similarity 21.8%; Pred. No. 1.1e-20;
Matches 149; Conservative 117; Mismatches 284; Indels 132; Gaps 25;

QY 8 RHLLVQLALLP-----AATQGNKYVGLGKGDVTELTCTAQQKSIOPHMNSNOIKI 61
DB 2 KHLMLFLVLAARWVLSQVLOQMGAGLKPSETSLTCTGCVGSGFSGYWMIKPPG 61
QY 62 LGNQ--GSFLTKGPSKLTNDRADSRSL-WDQGNFPLIKNLKIEDSDT--YICEVEDQKE 116
DB 62 KGLEWIGEINHSSTWNPBLKSRVITISVDTSKQLSLKSSVNAADTAVYYC----- 114
QY 117 EVQLVFGLTANSDTHL--LQGSLLTLESPPGSSP-----VQCRSPRKNIQGGKTL 169
DB 115 --ARVITRAPSGTDGRGYMDVWGCTTVTVSSGSASAPLTPLVSC----- 158
QY 170 SVSGLQELDQSGMTCTVQNGKQVEFKIDIVLAFQ-KASSIVKKEGEVFPPLAFT 228

```

Db 159 -----ENSPDSTSVAVGCLAQDPLPSITPSPMKYKNNSDISSTRG-----PFSVLAR 205  
 QY 229 VEKLTSSGGLMMQAEKASSSKMTTDLKKNKESYKRVYTQDPLQMGKLLPLHL--TLFQ 286  
 Db 206 GGYKATTSQVLLPSKGV-----MOGTDEHVCKVQHPNGKKNKXVPLPVIAELRP 255  
 QY 287 ALPQVA-----GSGNLTALAEAKTGKHOEYNLVNMR-----ATOLQKNTCEVM 331  
 Db 256 KSVFPPPPDGGFFGNPKRSKLTICATGSPSPQIQVMIWLEGKQVSGVTTDVOQAFAKS 315  
 QY 332 GPTSPKMLMLKLENKAKVSKREKPPVVLNPEAGMOCILSDSGVLLP-SNIXYLPW 390  
 Db 316 GPFTYKVTSTLTIKESD-----ML--SQSNFTCRVDHRLGLTPQNNASSMVCVPQ 362  
 QY 391 STFPVCPAPBPYKSCDXTHTCPPELLGSPSVFLPPPKKDTLMISRTREYTCVVVDVSHRP 450  
 Db 363 DTAI-----RVFAIPPS-FASIFLTGSKTCLTCLVTLDTLTYD- 397  
 QY 451 EYKFNMYVDGVEVHNAKTKPREQYNSTYRVVSVLTVLHQDMLNGKREYKCKNSKALPAP 510  
 Db 398 SVTISSTRONGEAVKTHNTISHPNATPSAVEASICEEDMNSGERFTCTVTHHTLPSRP 457  
 QY 511 IEKTIKAKGQP-REFQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQP-- 566  
 Db 458 LKQTIIRPKGVALHRPDDVILLPPARQQLNRESATITCLVTGSPADVFVQMMQKQPLS 517  
 QY 567 ENNYKTPPYLD--SDGSFPLYSKLTVDKSRWQGVFSGVNHKLNHYTKSLSP 624  
 Db 518 PEKYVASAPPEPQAGRYFAHSILTVSEEMNTGETYTCVVAHEALPNRIVERTYDKS- 576  
 QY 625 GLQDTCAGADQDGLDGLMTT 646  
 Db 577 -----TEGEVSADEEGFENLMAT 594

## RESULT 46

Q8VCX7 PRELIMINARY; PRT; 613 AA.  
 AC Q8VCX7;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS IGH-6.  
 GN Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Salivary gland;  
 RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC018315; AAH18315.1; -  
 DR MGD; MGI:96448; IGH-6.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 5.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 5.  
 DR PROSITE; PSS00290; IG\_MHC; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 613 AA; 67855 MW; 41A9384DD4C22862 CRC64;

Query Match 11.1%; Score 380.5; DB 11; Length 613;

Best Local Similarity 22.5%; Pred. No. 1.4e-20;

Matches 157; Conservative 97; Mismatches 226; Indels 217; Gaps 28;

QY 30 LGKKGDVELTCTAS-----QKSIQFHWKNSNQIKILGNQS--FLTKGPS 74  
 Db 30 LMKPGASVKISKATGVTFSYWIEMVQKPGHGLEWIG---ELIPGSSVYNEKFKG 85

QY 75 KLNRDAD--SRSLMDQGNFLLIKKLTIEDSDTYICEVEDQKEVOL-----VFGTLTA 127  
 Db 86 KATFTADTSSNANVQ-----LSSLTSEDSAVYIC-----ARLRGRWYFDWG--- 128  
 QY 128 NSDTHLQGGSTLTLESF--PGSSPSYQCRSP-RGKNIQGGKTLSSVSOLEIQQSGTWC 184  
 Db 129 -----AGTVTVSSESSQSPFNVPFLVSCBEPPLSDKXILVAMGCLARLPFLPSTISFTWN- 180  
 QY 185 TVLQNGKVEFKIDIVLAFQKASSIVYKKESEQVFFSPPLAFYBKLTGSELMMQAR 244  
 Db 181 -----YQNTETVQ----- 189  
 QY 245 ASSSKWITFDLKNKESYKRVYTQDPLQM-----GKPLHLTLPLQALPQ 290  
 Db 190 --GIFTFLRKGGKYLATSQVLSFKATLBESDEYLVCKINHYGGNRLHPIPI----- 242  
 QY 291 YAGSNLTALAEAKTGKHOEYNLV-----MRATQLOKNTLCEVMGFTSPKMLSLKL 344  
 Db 243 -----AVAMNPVNVFVPPRDPGSPAPRKSCLICEATNFTPPRITVSMK 289  
 QY 345 ENKEAKVSKREKPPVVLN--PEA-----GMQCLSDSGOVLLSESN 383  
 Db 290 DKLVESEGTTPDVTIEKNGSTPQYKVISLTLSIEIDMLNINVTYCRVDHRLGTLFK-- 347  
 QY 384 IKVLPWTSTPVPCPAPBPYKSCDXTHTCPPELLGSPS--VFLPPPKKDTLMISRTREY 439  
 Db 348 -----NVSSTC--ASPSIDILFTIIPSPAD-IFLSASANT 382  
 QY 440 CTVVDVSHEDPEVKENMYVDGVEVHNAKTKPREQYNSTYRVVSVLTVLHQDMLNGKEYK 499  
 Db 383 CLVSNLATYE-TLNTISMASQSGEPLETYKIKIMESHNPNGFSAKGVASVCVEBMNRKEV 441  
 QY 500 CYSNKAALPAPIEKTISKAKGPRE-----POVYTLPPSRDELTKNQVSLTCLVKGF 553  
 Db 442 CTVTRHDPSPQKFKISK--PNEVHKRPAPVYLPPARQQLNRESATITCLVAGESP 497  
 QY 554 SDIAVEMESNQ--PENNYKTPPYLD--SDGSFPLYSKLTVDKSRWQGVFSGVNHKLNHYTKSLSP 609  
 Db 498 ADISVQMLRGQLLQEKYVTSAPMPBPAPGFYFTHSLITYTEEMNSGERFTCTVGVH 557  
 QY 610 ALNHYTKSLSPGLQDTCAGADQDGLDGLMTT 646  
 Db 558 ALPHLVTERTVDKS-----TEGEVNAEEGFEENLMAT 589

## RESULT 47

Q9BU10 PRELIMINARY; PRT; 597 AA.  
 AC Q9BU10;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymph;  
 RA Strausberg R.;  
 RL Submitted (FEB-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC002863; AAH02863.1; -  
 DR HSSP; P01825; 7FAB.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR InterPro; IPR003596; Ig\_v.  
 DR Pfam; PF00047; Ig; 5.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 5.  
 DR PROSITE; PSS00290; IG\_MHC; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 597 AA; 65274 MW; 2DAFA8FB7E055851 CRC64;



OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymph;  
 RA Strauberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC019235; AAI19235.1; -  
 DR InterPro; IPR007110; IG\_Like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; IG; 5.  
 DR SMART; SM00406; IG; 1.  
 DR PROSITE; PSS0835; IG\_Like; 5.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 KW Hypothetical protein.  
 SQ SEQUENCE 588 AA; 64438 MW; FC60DBAD82B39FD7 CRC64;  
 Query Match 10.9%; Score 372.5; DB 4; Length 588;  
 Best Local Similarity 21.9%; Pred. No. 5.6e-20;  
 Matches 144; Conservative 115; Mismatches 273; Indels 127; Gaps 24;  
 QY 8 RHLLVQLALP-----AATGKNVVGKGGDTVELCTASQKKSIOFHMKNSNOIKI 61  
 DB 2 KHLMPFLVLAARWVLSQVQLOMGAGLIKSEITSLTGVYGGSEFGYMGWITQPPG 61  
 QY 62 LGNQ--GSFLTPKPSKINDRADRSRL--WDGPNFLIKKLIEDSDT--YICEVEDQKE 116  
 DB 62 KGLEWIGEINHSGSTVNPFLSKRSVITSDTSKKQLSLKSVNADTAVYC----- 114  
 QY 117 EVLLVFGULANDTHL---LOGSLTLTLESFPGSSPS---VQCRSPKNGIKQSGKTL 169  
 DB 115 --ARVITRAPPGIDRGVMDVWGQTTVIVSSGSAAPLPLVSC----- 158  
 QY 170 SVSQLELDGSGTWTCTVLQKQKVEFKIDIVLAFQ--KASSIYKKEGEVEFPLAFT 228  
 DB 159 -----ENSPDSTSSAVAGCLAQDFLPDSITSMWKKNSSDLSIRG-----FESVLR 205  
 QY 229 VEKLTSGSELWQAEPRASSKSWITPDLKNEKVSVKRVTPQPKLQNGKKPLHL--TLQP 286  
 DB 206 GGRYAATSGVLPSPKDV-----MQCTDEHVVCVKVQHPNGNKEKNVPLVIAELRP 255  
 QY 287 ALPQYA-----GSGULTLALFAKTKLHGVNVLVNR-----ATQLQKNLTCVW 331  
 DB 256 KVSVPFPPRDGFGFNPKSKLIGQATGFSRPIQVSWLRGKQVSGGVTTDQVQAEAKS 315  
 QY 332 GPTSPKLMISLKLKNEKAKVSKREKPVWVNLNPAQMMQCLSDSGQVLE--SNIKVLPW 390  
 DB 316 GPTTYKVTSLTLTKESD-----WL---SQSMFTCVDRHGLTFCQNASMGCVPDQ 362  
 QY 391 STFPVCPAPPPKSCDKTHTCPPELLGSPVFLFPKPKKDTLMISRTPEVTCVVVDVSHEDP 450  
 DB 363 DTAI-----RVPAIPPS--FASIFLTKSTLCTLVLDLTLYD-- 397  
 QY 451 EYKFNMYVDGVEVNAKTKRREQVNSTYRVVSVLVLYLHODWLNKGEYKCKKSNKALPAP 510  
 DB 398 SVTISWTRONGEAVKTHNISESHNPATFSAVEBASICEQNMNSGRFTCTVHTDLPSP 457  
 QY 511 IEKTIKAKQGP--REPOVYTLPPSRDELTL--KNQVSLTCLVKGYFSDIAVEMESNGOP-- 566  
 DB 458 LKQTTISRKQVALHRRPVYLLPPAREQLNRESATITCLVTGSPADVPQVMQROQPLS 517  
 QY 567 ENNYKTTTPVLD--SDGSEFLYSKLTIVDKSRMOQGVNFCVSWHEALHNHTYKSLSLS 623  
 DB 518 PEKYVTSAPWPEQAPGRYFAHSILTVSEEMNTGFEYTCVVAHEALPNRVTERTYDKS 576

DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CD4 protein (Fragment).  
 GN CD4.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Zverev V.V., Blinov V.M., Nedospasov S.A.;  
 RT "Splice-mediated insertion of antisense and sense Alu repeats in human CD4 gene: identification of three exons of CD4 mRNA."  
 RL Submitted (APR-1993) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE; 95407135; PubMed=7676667;  
 RA Zverev V.V., Sidorov A.V., Nedospasov S.A., Maliushova V.V.,  
 RA Udolova I.A., Andzhaparidze O.G., Blinov V.M.;  
 RT "[Nucleotide sequence of two exons of the human T-lymphocyte CD4 receptor gene]."  
 RL Vopr. Virusol. 40:100-102 (1995).  
 DR EMBL; X87579; CAA60883.1; -  
 DR EMBL; S79267; AAB35273.1; -  
 DR PIR; I60082; I60082.  
 DR HSSP; P01730; 1CDY.  
 DR GO; GO:004872; F:receptor activity; IEA.  
 DR InterPro; IPR007110; IG\_Like.  
 DR PROSITE; PSS0835; IG\_Like; 1.  
 KW Receptor.  
 FT NON TER 71 71  
 SQ SEQUENCE 71 AA; 7844 MW; A5C9D84816135C66 CRC64;  
 Query Match 10.5%; Score 357; DB 4; Length 71;  
 Best Local Similarity 98.6%; Pred. No. 4.2e-20;  
 Matches 70; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 MNRGVPRHLLVQLALPAAQGNKVVYLGKGGDTVELCTASQKKSIOFHMKNSNOIKI 60  
 DB 1 MNRGVPRHLLVQLALPAAQGNKVVYLGKGGDTVELCTASQKKSIOFHMKNSNOIKI 60  
 QY 61 ILNGGSFLLTK 71  
 DB 61 ILNGGSFLLTK 71  
 RESULT 51  
 ID 086TT1 PRELIMINARY; PRT; 375 AA.  
 AC 086TT1;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Human full-length cDNA clone CS0DD006YL02 of neuroblastoma of Homo DE sapiens (Human).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neuroblastoma;  
 RA Genoscope;  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neuroblastoma;  
 RA Li W.B., Gruber C., Jessee J., Polayes D.;  
 RT "Full-length cDNA libraries and normalization."  
 RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BX161420; CAD61894.1; -  
 DR GO; GO:0046821; C:extrachromosomal DNA; IEA.  
 DR InterPro; IPR007110; IG\_Like.



```

RL Dev. Dyn. 225:384-391(2002).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=2238257; PubMed=12477932;
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schler G.D.,
  Hopkne R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
  Diatchenko L., Martins K., Farmer A.A., Rubin G.M., Hong L.,
  Stapleton M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
  Brownstein M.J., Uddin T.B., Tohyuki S., Carninci P., Prange C.,
  Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Miliady S.J.,
  Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
  Richardson S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
  Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
  Fahy J., Helton E., Kesteman M., Madan A., Rodriguez S., Sanchez A.,
  Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
  Rodriguez R.W., Touchman J.W., Green E.D., Dickson M.C.,
  Krzyzinski M.I., Skalska U., Smalms D.E., Schermer A., Schein J.E.,
  Jones S.J., Marra M.A.;
  "Generation and initial analysis of more than 15,000 full-length human
  and mouse cDNA sequences."
RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN [3]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RX Klein S., Strausberg R.;
  Submitted (Aug-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC056078; AAH56078.1; -.
KW Hypothetical protein.
SQ
SEQUENCE 587 AA; 64933 MW; 64635FD0FA148D CRC64;

Query Match
Best Local Similarity 9.5%; Score 324.5; DB 13; Length 587;
Matches 98; Conservative 54; Mismatches 115; Indels 65; Gaps 14;

QY 342 LKLEKAKVSKREKV-----WLNPEAGMOCIL--SDSQVLL 380
DB 280 LKNGNQTGEGVVEEPEDEKRGYEATSYLSTRKEMDLD--LVSCVEHAESES-LQ 335
QY 381 ESKIKVLPWSTVPAPAPKSCDKHTPCRLGGSVFLPPKPKDITLMTSRTEBVC 440
DB 336 EKMM-----SKSLMCDTPPTPTSTIOTITP-----PSL-----ESTFEKKSATLTC 376
QY 441 VVVDVSHEDPEVKFNNY--VDGVEVHNAKTPREEOYNS--TYRVSVLTVLHODMNGKE 497
DB 377 LVSNMANSBDRISINPKKSGTQEIPLKTELGDATINDNRTYSVKGTTVCADENNDK- 435
QY 498 YKCKVSNKALPAIEKTIKAKQCPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDI 556
DB 436 FVCKVETELASKEVFLFKKEGNYTPSYVPPPELSEKREKATLTCLTVKGFSPSEI 495
QY 557 AVEM--ESNGCPENNYKTP-----PVLDSDGSPFLYSKLTVDKSRMOQGNVFSGVNH 608
DB 496 FVEMHNGEAVPKONTYINTSINDELLPKGQKSGKFFLYSLHTIDIDMDAGDSFCVGH 555
QY 609 EALNHVYTKSLSPG-----LQLDCTC 632
DB 556 ESLPLQLTKRSIDKSSGKPTNNVNSLVLSLSDTC 587

RESULT 54
Q27379 PRELIMINARY; PRT; 478 AA.
AC 072379;
DT 01-OCT-2003 (TEMBLrel. 25, Created)
DT 01-OCT-2003 (TEMBLrel. 25, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE Hypothetical protein DKFZp686K04218 (Fragment).
GN DKFZp686K04218.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=human rectum tumor;
RA Bloecher H., Boeher M., Mewes H.W., Weill B., Amid C., Oeanger A.,
  Rolo G., Han M., Wiemann S.;
  Submitted (JUN-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL, BX538066; CAD9796.1; -.
KW Hypothetical protein.
FT NON TER
SQ
SEQUENCE 478 AA; 51620 MW; 4AFCB541F3217CA1 CRC64;

Query Match
Best Local Similarity 9.2%; Score 313.5; DB 4; Length 478;
Matches 131; Conservative 65; Mismatches 177; Indels 177; Gaps 25;

QY 171 VSQLELDPSG-----TWCTVLQNKKEFKIDIVLAFQKASSIVYKKEGEQVE 220
DB 17 LSQVQLQESGPELVKPSQTLSTCTV-----SGSI----- 47
QY 221 FSPPLAFTVEKLTGSGELMWOAERASSSKSWITFDLKNKEVSKVTDPKLQMKUPL 280
DB 48 -----GSGDYFW-----SWI-----RQAP-----GRGLEW 67
QY 281 HLTPLQALPQYAGS-----GNITLAEATGKLGHOEVN-----LYVNRATQLOKN 325
DB 68 -----MGYIYSGSYTYNPSLESRLSISIDSKNFKSLNSLTADTAIVYFCAGVGIG 122
QY 326 LTCEVWG-----PTSPKLM-LSLKLEKAKVSKR-----EKPVWVLANPEA 365
DB 123 TAFDWMGGQVTVTVSSASPTSPKFPPLSLDSTFPQGNVYVACLQVGFQPELSTWSSS 182
QY 366 GN-----WQCLSDSGOVLLSNIKVLPTWSTP-----VPCAP 399
DB 183 GGNVTARNPPEQDASGDLTYTSSQTLTPATCPDGKSVTCVHKHYTNPSQDVTPCVP 242
QY 400 EPKSCDKHTPCRLGGSVFLPPKPKDITLMTSRTEBVC-----VPCAP 459
DB 243 PPPPC-----C-----HPLSLHRALEED-LLLGEANLTCTLTGL-RASGATFTWPS 290
QY 460 GVEVHNAKTPREEOYNSYRVSVLTVLHODMNGKEKCVSNKALPAIEKTIKAK 519
DB 291 SGK--SAYGPPERDLGCTSVSVLPQCAQGMNGETTCTAAHDELKTLPLANTTKS- 347
QY 520 GQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNGQ--PENNYKTPPV 576
DB 348 GNTFPEPVHLLPPSEELALNELVTLTCLARGFSPDVLVRLQSGQSELPREKYLTMASR 407
QY 577 LD-SQG--SFLYSKLTVDKSRMOQGNVFSGVMEALNHVYTKSLSPG----- 625
DB 408 QPSQGTTFVATYSILRAAEDWKGDFTSCVSHSEALPLAFTQKIDRLAGKPTHVNS 467
QY 626 ---LQLDCTC 632
DB 468 VYMAEYDGTIC 477

RESULT 55
Q29027 PRELIMINARY; PRT; 99 AA.
AC 029027;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TEMBLrel. 25, Last annotation update)
DE CD4, allele 1 (Fragment).
OS Sus scrofa (Pig)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.

```



```
RC STRAIN=1183; TISSUE=Blood;
RA MEDLINE=93329116; PubMed=8335933;
RT "Guatason K.; Germana S.; Sundt T.M.; Sachs D.H.; Leguern C.;
RT "Extensive allelic polymorphism in an exposed region of the
RL J. Immunol. 151:1365-1370(1993).
DR EMBL, X65629; CAA46583.1; -.
DR PIR, I47131; S21461.
DR HSSP, P01730; 1CDY.
DR InterPro: IPR007110; IG_1-like.
DR InterPro: IPR003596; IG_v.
DR Pfam: PF00047; IG_1.
DR SMART, SM00406; IGV_1.
DR PROSITE, PS50835; IG_LIKE; 1.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 99 AA; 11170 MW; 40BF080699CF5D0C CRC64;

Query Match 8.9%; Score 305.5; DB 6; Length 99;
Best Local Similarity 60.2%; Pred. No. 6,1e-16;
Matches 59; Conservative 16; Mismatches 22; Indels 1; Gaps 1;

Qy 32 KKGDTVELTCTASOKSIQEFHMKNSQIKIINGQSF--TKGPSKLNDRADRSRLNDQG 90
Db 1 KAGDLAELPCRSSQKKNLPFWMKNSQTKILGGHGFMTASVYELTSRLDSKKNMMDHG 60

Qy 91 NFPLIINKLKIEDSDTYICEVEDQKEEYQLVFGLTAN 128
Db 61 SFPLIINKLEVTDSGIYCEVEDKRIEYQLVFLRLTAS 98

RESULT 56
Q9NPP6 PRELIMINARY; PRT; 416 AA.
AC Q9NPP6:
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Immunoglobulin heavy chain variant (Fragment).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Suffert R., Estivill X., Escarceller M., Sunoy L.;
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Auffray C., Anorge W., Ballabio A., Estivill X., Gibson K.,
RA Lehrach H., Pouletka A., Lundberg J.;
RT "The European IMAGE consortium for integrated molecular analysis of
RT human gene transcripts.";
RL Submitted (JUL-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL, AL389978; CAB97534.1; -.
DR HSSP, P01789; IMCP.
DR InterPro: IPR007110; IG_1-like.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_v.
DR Pfam, PF00047; IGV_1.
DR SMART, SM00406; IGV_1.
DR PROSITE, PS50835; IG_LIKE; 3.
DR PROSITE, PS00290; IG_MHC; 2.
FT NON_TER 1
FT NON_TER 99
SQ SEQUENCE 416 AA; 44786 MW; 8C41708BB8AB4687 CRC64;

Query Match 8.8%; Score 302; DB 4; Length 416;
Best Local Similarity 27.8%; Pred. No. 8,9e-15;
Matches 102; Conservative 47; Mismatches 140; Indels 78; Gaps 16;

Qy 329 EWNG-----PTSPKLM-LSLKLKNEKAVSKR-----EKPVVNLNPEAGM- 367
Db 64 DWNGQTTTVTSASPTSPKVFPLSLDSTPDGNNVVVACLVGFFPPEPLSVTMSSEGN 123
```

```
Qy 368 -----MOCLSDSGQVLLIESNIKVLPTWSTP-----VPCPADEPK 402
Db 124 VTANFPSPQASDGLVTTSSQLPLPANCQPDGKSVCHVGHYTNPSQDVTVPCEVPPR 183

Qy 403 SCDTHTCPPELLGSPSFLPPPKXTLMTSRPEVTGVVVDVSHEDPEVKNYVDGVE 462
Db 184 PC-----C-----HPRLSLHRPALED--LLIGSEANLCTTLTGL--RDASGATFTWPSGK 231

Qy 463 VHNKTPREKQVNSTRVVSVLTVLHODMNGEYKCNKALPAPIETISKAGQP 522
Db 232 --SAVQBPPEKDLCCYSVSSVLPFGCAQPMNHGFTFTCTAAHPELKTPTANITKS--GNT 288

Qy 523 REPQVYTLPPSRDELTKNQ--VSLTCLVKGFYPSDIAVEMESNQ--PENNYKTTPLVD- 578
Db 289 FRPETHLPPESELANELVTLTCLARGSPKQDVLRVNIQSGSELPREKILYTWASQEP 348

Qy 579 SDG--SFFLYSKLTVDKSRMQGNVFSQVMEHALNHNHTQKSLSLSPG----- 625
Db 349 SQGTTTAVVSIILVAEDMKKGDTFCGVCVGHFALPLAFQTKTIDRLAGKPTHVNVSM 408

Qy 626 LQDDETQ 632
Db 409 AEVDGTC 415

RESULT 57
Q96K68 PRELIMINARY; PRT; 494 AA.
AC Q96K68:
DT 01-DEC-2001 (TREMBlrel. 19, Created)
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Hypothetical protein FLJ14473.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isogai T., Ota T., Hayashi K., Sugiyama T., Otsuki T., Suzuki Y.,
RA Nishikawa T., Nagai K., Sugano S., Shiratori A., Sudo H.,
RA Magatsuma M., Hosoi T., Kaku Y., Kodaira H., Kondo H., Sugawara M.,
RA Takahashi M., Chiba Y., Ishida S., Murakawa K., Ono Y., Takiguchi S.,
RA Watanabe S., Kamura K., Murakami K., Ishi S., Kawai Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Nagahara K., Maeno Y.,
RA Niinomiya K., Iwayanagi T.;
RT "NDO human cDNA sequencing project.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, AK027379; BAB55072.1; -.
DR PIR, S21205; S21205.
DR InterPro: IPR007110; IG_1-like.
DR InterPro: IPR003006; IG_MHC.
DR InterPro: IPR003596; IG_v.
DR Pfam, PF00047; IGV_1.
DR SMART, SM00406; IGV_1.
DR PROSITE, PS50835; IG_LIKE; 4.
DR PROSITE, PS00290; IG_MHC; 1.
KM Hypothetical protein.
SQ SEQUENCE 494 AA; 53088 MW; 9A1D7A85A8E4C0E CRC64;

Query Match 8.8%; Score 300.5; DB 4; Length 494;
Best Local Similarity 21.8%; Pred. No. 1,5e-14;
Matches 139; Conservative 82; Mismatches 208; Indels 209; Gaps 27;

Qy 30 LGKKGDTVELTCTAS--QKSIQFHW-----KNSQIKIINGQSF-----TKPSKL 76
Db 30 LVKPGSLRLSCAASGSPSTYAMNVRQAPGKGLVWVISISSRSDIYYRDSVKGRFTI 89

Qy 77 NDPAADRSRLMDGNNPPLIINKLKIEDSDTYICEVEDQKEEYQLVFGLTANSDTLQD 136
Db 90 S-RDNKNSISLYQMN-----SLRVDDTAVTYICARDSCNGAI--CYGFSF-----WG 132
```

```

Oy 133 QSLTLTLESPGSSSSVQCRSPRGKNIQGGKTLASVQLELDQSGWTCVTVLQNGKVEFK 196
Db 133 QGLTVTSSASPTSP-----KVFLSLCSTQPDG-----NVV 164
Oy 197 IDIVLVAE--KASSIVYKKEGEQVEF-SPLAFLVEKLTGSGELMMOAEKSSKSWIT 253
Db 165 IACLVGFFPQEPPLSVTWSESGVTANNF-----PQDASGDLY-----TTSQ----- 210
Oy 254 FDLNKEVSVKRVITDPKLQNGKKLPLHLTFQALPQYAGSGLTLALEAKTGKLDQEVN 313
Db 211 -----LTLF----- 214
Oy 314 LVWRARQ--LQKLTCEVWGPTSPKMLSLKLENKEAKVSRERKPVWVNLNEAGMQCL 371
Db 215 -----ATQCLAGKSYTCHVKHYTNP----- 234
Oy 372 LSDSGQVLLESNIKYLPMTWSTVPQAPAPKSCDKTHTCELLGGPSVFLPPKPKDTLM 431
Db 235 ---SDQVTVPCVPSTPPTPSPSTPPTPSPSCCH-----PRLSLHNPALFD--LL 279
Oy 432 ISRPDEVTCVVVDVSHEDPEVKFNKYVDGVEVHNAKTKRREQYNSTYRVVSVLTVHLQD 491
Db 280 LGSSEANLTCTLTGL-RDASGVFTFWTPSSGK--SAVQGPENIDLGCYSVSVLPGCAEP 336
Oy 492 WINGEVYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPSRDELTKNQ-VSLTCLYKG 550
Db 337 WNHGTFCTCTAAVPESKTPPLTATLTKS--GNTRPREVHLTPPSEETALNETVLTITCLARG 395
Oy 551 FYPSPIAVEWESNGQ--PENNYKTTTPVLV--SDG--SFLYSKLTLYDSRMQGGVFSQS 605
Db 396 FSPKVLRLRWLGSGSELPREKLTWASROEPOGTTTFAVISILRLVAEDMKKGDTFSCM 455
Oy 606 VMHEALHNHYTQKSLSLSPG-----LQLEDTG 632
Db 456 VGHEALPLAFTQKTIIDLAKGRTHVNVSVMAEVDGTC 493

RESULT 58
O96KX8
O96KX8 ID O96KX8 PRELIMINARY; PRT; 496 AA.
AC O96KX8;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
RL Struhsberg R.;
RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC016369; AAH16369.1; -.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_V.
DR Pfam; PF00047; IG_4.
DR SMART; SM00406; IGV_1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PSS00290; IG_MHC; 1.
DR Hypothetical protein.
SQ SEQUENCE 496 AA; 53391 MW; D346923849040D69 CRC64;

```

Query Match	8.7%	Score 298	DB 4	Length 496
-------------	------	-----------	------	------------

Best Local Similarity 23.3%; Freq. NO. 2.3e-14;  
Matches 132; Conservative 69; Mismatches 209; Indels 156; Gaps 21;

QY 124 GLTANSDTHLLQGSGITLTLESPPGSSPSVC---CPSPRGNIGGKTLISVQL-----E 175  
 Db 29 GLVKSEET-----LSLTCTVAGSGSISSSYWGMWIPPGKGLEWITANTYYSGITYNPS 83

```

QY 176 LODGSTMTCYLOMKGKVEFKIDIVLAFOKKASJIVYKKEGQVESFELAFVEKLTGS 235
Db 84 LKSRNTISVDRISKQ-----LSLKVRSVTJADTAIVY-----FCANHGVSRSRGRTGA 129
QY 236 GELWMOAERASSSKSWITFDLKNKEVSVKRTQDPKLOMKGKLPJLHJLTPALPOYAGSG 295
Db 130 IDYMOGQGLTVTVSSASPT-----SPKV-----FPLSLCSTOP-----DG 163
QY 296 NULTAL-----EAKTGKLDHOEVLVVMARATQ--LQ 323
Db 164 NVVIAICLVQGFPPQEPBESLVTMSSEGGVYARNFPPSQDASGDLVYTTSSQGLTPATQCLAG 223
QY 324 KNLTCVWGPTSPKJMLSLKLENKEAKVSKREKPVWVLNPEAGMOCCLSDSQVLLEBN 383
Db 224 KSVTCHVGHYINP-----SODVTVPCP 245
QY 384 IKVJPTWSTPVPCPAPBPKSCDKHTTCPELLGSPSVLEPRPKQDTLMSIRPEVVCVY 443
Db 246 VPSTPPTSPSTPPPPSPSCCH-----FRLSLHRPLED--LLGSEANLITCTLT 233
QY 444 DVSHEDPEVKENWYDVGEVHNAKTPKREBOYNSYTVRVSVJLTVLHQDMLNGKEYCCKVS 503
Db 294 GL-RDASGVTFWTWPSGK--SAYGQRPREDLCGCSVSASYVPGCAEPYNNHGTFTCTAA 350
QY 504 NKALBAPLEKTIKSAKGQRPBOVYVTLPRSDBELTKNQ--VSLTCLVKGSPVSDIAEWBS 562
Db 351 YPESKTPJLTATLSKS-GNTPREPVHLLPRPSEELALNELVTLTCLANGSPKQVLAWRMLQ 409
QY 563 NGQ--PENNYKTPVLVD-SDG--SFPLYSKJLTVDSRMQOQNVFCSVMHEALHNHYTO 617
Db 410 GSQELPREKXYLTWASRGQPSGQTTFAVATVSLRVAADWKKGDTFSCMGHEALPLAFTQ 469
QY 618 KSLSLSPG-----LQDLENC 632
Db 470 KTDRLACKPHTHVNVSVMAEVDGTC 495

```

**RESULT 59**

Q72374 PRELIMINARY; PRT; 492 AA

DT 01-OCT-2003 (TREMBlrel. 25, Created)

DT	01-OCT-2003	(TReMBLrel. 25, Last annotation update
----	-------------	--

DE hypocoelomic process (fragment)  
GN DKFZP686C02218.

US Homo sapiens (Human) ;  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea;  
 OX NCBI TaxID=9606;

RN	[1]	SEQUENCE FROM N A
BP		

RC	TISSUE=Human rectum tumor;		
PA	Boecker H	Mewes H W	Weil B
			Amid C
			Osander A

RA Fobo G., Han M., Wiemann S.;  
 submitted (PM-2002) to the EMBL/GenBank/DBJ databases

DR EMBL; BX538077; CAD98001.1; -.

FT	NON TER	1	1	15331577C0E0C37AB CBCE4
00	STATION	400	11	15331577C0E0C37AB CBCE4

1. **Introduction**

Best Local Similarity 27.7%; Pred. No. 2.7e-14;

[illegible]

27

DD 14Z M06GILV1VSDH0F10FNVFDDDD01F000NVVWCLV0011Z00LEDD11NOED0011Z00

368 ---WUCLBSBSGVLEBSNINAVPIMSI P-----VCLCESEFIWOC 700

DB 202 ARNFPPS QDASG DLTITISS Q LILFAIQCPDGSVICHVNH I NPSQDVI VECFVEFFEEC 203

```

Qy 405 DKHTCEBLGPGPVFLPPPKKDTLMISRTPEVTCVVDVSHEDPEVKFNMVYDGEVH 464
Db 262 -----C-----HPLSLHRLPALD-LILGSEANLCTLTGL-RASGATFTWTSBGK-- 307
Qy 465 NAKTRPEEQYNSTYRVSVLTVLHQLMNGKEYCKVSNKALPAPIEKTISKAGOPRE 524
Db 308 SAVGQPPERDLGGCVSVSLPGCAQPTWGTETCTAAHPELKTPLTANITKS-GNTFR 366
Qy 525 POYVTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEESNQ--PENNYKTPPLVD-SD 580
Db 367 PEVHLPPSEELALNELVTLTCLARGSPKDVLRWLQSGQLPREKYLTMASRQEPQ 426
Qy 581 G--SFFLYSKLTVDSKRMQGNFSCSYMEALHNHYTKSLSPG-----LQ 627
Db 427 GTTFATYSLIRVAEDMKKGDTSQVGHFLPLAFTQKTIIDLAKGPTHVNVSVMAE 486
Qy 628 LDFTC 632
Db 487 VDGTC 491

```

## RESULT 60

```

Q8NSK4 PRELIMINARY; PRT; 499 AA.
ID ID Q8NSK4
AC Q8NSK4;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RA Strauberg R.;
RL EMBL; BC032249; AA032249.1; -.
DR EMBL; BC032249; AA032249.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00409; IG; 4.
DR SMART; SM00407; IGcl; 2.
DR SMART; SM00406; IGv; 1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 499 AA; 53376 MW; 93A5C89582054F32 CRC64;

```

Query Match 8.7%; Score 296; DB 4; Length 499;

Best Local Similarity 25.2%; Pred. No. 3.3e-14; Indels 128; Gaps 24;

```

Matches 131; Conservative 60; Mismatches 201; Indels 128; Gaps 24;
Qy 219 VESFPLAFTVEKLTG-----SGELMWAQERASSKSWTFDLAKKEV 261
Db 1 MERGLSVPLVAILKGVOCEVLVESGGVVRPGSL-----RLSCATSGFTP---DS 51
Qy 262 SVKRVYDDPKLQMGKLP--LHLTLPOALPOYAGS--GNLTALAEATKGLHDEVNLVVM 317
Db 52 GASVVRQAP-----GKGLEWVSSIMNGSGSTVADSVGRFTISRDNKNSLYIQMNSLRV 107
Qy 318 RATOLQ-----KNLTC-----EVMG-----PNSPKLM--LSLKLKNEAK 350
Db 108 EDTALYICANDPTYYCGSGCLGYMYVMWGKTTVTVSASPTSPKVPFLSLCSTQPDGN 167
Qy 351 VSKR-----EKPVVNLPEAGM-----WQCLLD 374
Db 168 VVLAICLVGFPFGPRLSVTWSSESGGVYANRFPSPQASGDLVYTTSSQTLPLPATQCLAG 227

```

```

Qy 375 SGQVLESNIK--VLPTWSTPVPQAPPEKSCDKTHCPCLG---PSVFLPPPKKDT 429
Db 228 S-----YCHVKAHTNNSQDVTVPDPSTPTSPSPSTPPSPSCCHPRSLHRLPALD- 282
Qy 430 LMSRPEVTCVVDVSHEDPEVKFNMVYDGEVHNAKTRPEEQYNSTYRVSVLTVLH 489
Db 283 LILGSEANLCTLTGL-RASGATFTWTSBGK--SAVGQPPERDLGGCVSVSLPGCA 339
Qy 490 QDMLNGKEYCKVSNKALPAPIEKTISKAGOPREPOYVTLPPSRDELTKNQ-VSLTCLV 548
Db 340 EPMNHGKTFCTAAYPESKTPLTATLTKS-GNTFRPEVHLPPSEELALNELVTLTCLA 398
Qy 549 KGFYPSDIAVEESNQ--PENNYKTPPLVD-SDG--SFFLYSKLTVDSKRMQGNFVS 603
Db 399 KGFSPKDVLRWLQSGQLPREKYLTMASRQEPQGTTFATYSLIRVAEDMKKGDTS 458
Qy 604 CSYMEALHNHYTKSLSPG-----LQDFTC 632
Db 459 CMVGHFLPLAFTQKTIIDLAKGPTHVNVSVMAEVDGTC 498

```

## RESULT 61

```

Q8NCL6 PRELIMINARY; PRT; 493 AA.
ID ID Q8NCL6
AC Q8NCL6;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein FLJ90170.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Mammary gland;
RA Isogai T., Ota T., Nishikawa T., Hayaishi K., Otsuki T., Sugiyama T.,
RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hiro Y., Saito K.,
RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,
RA Masuno Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
RT "NEDO human cDNA sequencing project.";
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AK074651; BA01114.1; -.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00409; IG; 4.
DR SMART; SM00407; IGcl; 2.
DR SMART; SM00406; IGv; 1.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 493 AA; 53224 MW; 12BCD7E094777101 CRC64;

```

Query Match 8.6%; Score 294; DB 4; Length 493;

Best Local Similarity 22.9%; Pred. No. 4.7e-14; Indels 212; Gaps 30;

```

Matches 147; Conservative 77; Mismatches 207; Indels 212; Gaps 30;
Qy 25 GNRVVLGKKGDVETLCTAS--OKKSIQFHWKNSNQ-----IKIIGNQ-----SFLTK 71
Db 27 GGGVVL--PGSRLRLCAASGFRFRDYDMHWVRQSGEGLEWALLIWTGCTKYVSDSVK 84
Qy 72 GPKSLNDRADSRSLMDQGNFPLIINKLTIKEDSDTYICEVEDQKEEVQLVFLGLTANSBT 131
Db 85 GRLTVS-RDNYKNLTY-----LEMKSIGEDTAAYYC-ARDQ-----GYAGYGVF-----D 128
Qy 132 HLLQGSILTLLESPPGSSPSVQGRPRGNIGCGTTLVSQLELDSDGTWCTCTVLQNGK 191

```

```

Db      129 HMGGLTVTVSSASP--TSP-----KVFPISLCSTDPD-----160
QY      192 KVEFKDIVLAF--QKASSIVYKKEGEQYF--SFPLAFVEXKLTSGGELMMQAKRASSS 248
Db      161 --NVVIAICLVQGFPEPPLSVTSSGQGVYARNFP-----PSQDASGDLV-----TSS 208
QY      249 KSMITFDLKNKEVSVKRVTDPKLQMGKPLRLHLLTLPQALPOYAGSGNLLALEAKTKL 308
Db      209 Q-----LTLF-----213
QY      309 HOEVNVLVWRATQ--LQKLLTCEVWGPTSPKMLSLKLENKAKVSKREKPVWVNLPEAG 366
Db      214 -----ATCLAGKSVTCVHKRYTNP-----233
QY      367 MRCCLSDSGQVLLSNIKVLPTWSTVPFCPAPBPSCDKTHTCPELLGSPVFLPPPK 426
Db      234 -----SODVVPVCPVSPPTPSPPTPSPSCCH-----PRLSLRPAL 274
QY      427 KOTLMTSRTEPVTCVVVDVSHEDPEVKFNWYVDGVEVHNAAKTPREQVNSTYRVISVLT 486
Db      275 ED--LLGSEANLCTLTGL--RDASGVTFWTTPSSGK--SAVQGPBERDLCCYSVSSVLP 330
QY      487 VLDQMLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELITNQ--VSLT 545
Db      331 GCAEPNHHGKTFCTAAYPESKTPPLTATLSK--GNTFRPEVHLPPRSEELALNELVTLT 389
QY      546 CLVKGFPSPDIAVWESNGQ--PENNYKTTPEVLD--SDG--SFLYSKLTVDKSRMQGN 600
Db      390 CLARGSPKDVLRWMLQSGSELPREKYLTVASRQEPSCQTTTFAVVISILVAEDWKGKD 449
QY      601 VFSCSVNHEALHNHYTQKSLSLSPG-----LQLEDET 632
Db      450 TFCGMGHEALPLAFTQKTIIDRLAGKPTHVNVSVMAEVDGTC 492

```

## RESULT 62

```

Q9BRV0 ID Q9BRV0 PRELIMINARY; PRT; 500 AA.
AC      01-JUN-2001 (TrEMBLrel. 17, Created)
DT      01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Hypothetical protein.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Prostate;
RA      Strauberg R.;
RL      Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; BC005951; AAH05951.1; -.
DR      HSSP; P01789; IMCP.
DR      InterPro; IPR007110; IG-like.
DR      InterPro; IPR003006; IG_MHC.
DR      InterPro; IPR003596; IG_v.
DR      Pfam; PF00047; Ig; 4.
DR      SMART; SM00406; Igv; 1.
DR      PROSITE; PSS0835; IG_LIKE; 4.
DR      PROSITE; PS00290; IG_MHC; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 500 AA; 54154 MW; 0A9BF43F2A3CC6D9 CRC64;

```

Query Match 8.5%; Score 289; DB 4; Length 500;

Best Local Similarity 22.6%; Pred. No. 1.2e-13;

Matches 136; Conservative 67; Mismatches 203; Indels 196; Gaps 26;

QY 118 VQLVLELT-ANDSTHLGQSLTLTSPSSSPSVQCSPPKGNIGGKTLVSQLEL 176

Db 7 ILFLVAAATGAQSQVHLVQGAQVMS---FGASVRSVSKT-----43

QY 177 QDSGTWTCTVLTQNKVKEFKDIVLAFQKASSIVYKKEGEQYFSPPLAFVEXKLTSGS 236

```

Db      44 --SG-----YAFHTYSLI-----WROAPGGG 63
QY      237 ELMMQAKRASSSKSWITFDLKNKEVSVKRVTDPKLQMGKPLRLHLLTLPQALPOYAGSGN 296
Db      64 LEW-----MGWISPSSDN-----TRFAKFF-----QGR 86
QY      297 LTLALEAKTKGHQF-----NLVWRATQLOKRLTC-----EWSG-----332
Db      87 VTLTDTSTSTVYMEIRSLRSDDTAYVYCARRYCSYSCNDYVYVWVGKGTTVVS 146
QY      333 ---PTSPKLM--LSLKENKAKVSKR-----EKPVWVNLPEAGM-----367
Db      147 SASPTSPKFPPLSLCSTQDPDGNVLAICLVQGFPEPPLSVTSSGQGVYARNFPSSQDA 206
QY      368 -----WQCLSDSGQVLLSNIK--VLPWSTPVPFCPAPBPSCDKTHTC 410
Db      207 SCDLYTTSQTLPLPAQCLAGKS---VTCVHYHYNPQSDVTVPCPVPSTPTPAPSTP 262
QY      411 PELLGG---PSVFLPPPKOTLMTSRTEPVTCVVVDVSHEDPEVKFNWYVDGVEVHNAAK 467
Db      263 PTPSPSCCHPRLSLRPALED--LLGSEANLCTLTGL--RDASGVTFWTTPSSGK--SAV 318
QY      468 TKPREQVNSTYRVISVLTVLHQMNGKEYKCKVSNKALPAPIEKTISKAKGQPREPOV 527
Db      319 QGPBRDLCCYSVSSVLSGCAEPNHHGKTFCTAAYPESKTPPLTATLSK--GNTFRPEV 377
QY      528 YTLPPSRDELITNQ--VSLTCLVKGFPSPDIAVWESNGQ--PENNYKTTPEVLD--SDG-- 581
Db      378 HLLPPRSEELALNELVTLTCLARGSPKDVLRWMLQSGSELPREKYLTVASRQEPSCQTT 437
QY      582 SFLYSKLTVDKSRMQGNVFSCSVNHEALHNHYTQKSLSLSPG-----LQLE 630
Db      438 TFAVVISILRVAEDWKGKDTFSCVMGHEALPLAFTQETIIDRLAGKPTHVNVSVMAEVDG 497
QY      631 TC 632
Db      498 TC 499

```

## RESULT 63

```

Q96DK0 ID Q96DK0 PRELIMINARY; PRT; 496 AA.
AC      01-DEC-2001 (TrEMBLrel. 19, Created)
DT      01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT      01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE      Hypothetical protein FLJ25298.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
NCBI_TaxID=9606;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Gastric mucosa;
RA      Ishibashi T., Kaneshori K., Yosida M., Watanabe S., Ishida S., Ono Y.,
RA      Hotta T., Hiraoka S., Murakawa K., Takiguchi S., Kusano J.,
RA      Watanabe M., Fujimori K., Tanai H., Ishida M., Yamashita H., Chiba Y.,
RA      Suzuki Y., Hata H., Nakagawa K., Mizuno S., Morinaga M., Kawamura M.,
RA      Sugiyama T., Irie R., Otsuki T., Sato H., Nishikawa T., Sugiyama A.,
RA      Kawakami B., Nagai K., Isogai T., Sugano S.;
RT      "NEO human cDNA sequencing project."
RL      Submitted (Oct-2001) to the EMBL/GenBank/DBJ databases.
DR      EMBL; AK058027; BAB71633.1; -.
DR      InterPro; IPR007110; IG-like.
DR      InterPro; IPR003006; IG_MHC.
DR      InterPro; IPR003596; IG_v.
DR      Pfam; PF00047; Ig; 4.
DR      SMART; SM00406; Igv; 1.
DR      PROSITE; PSS0835; IG_LIKE; 4.
DR      PROSITE; PS00290; IG_MHC; 1.
KW      Hypothetical protein.
SQ      SEQUENCE 496 AA; 53532 MW; C72E1E1247C86FED CRC64;

```

Query Match 8.5%; Score 288.5; DB 4; Length 496;  
 Best Local Similarity 22.7%; Pred. No. 1.2e-13;  
 Matches 134; Conservative 65; Mismatches 192; Indels 199; Gaps 26;

QY 129 SDTHLLQ-GGSLTLTLESPSSPSVQCRSGKNIGGKTLVSQLELDGSGTWTCTVL 187  
 19 SQVHLVOSGAEIKM-----PGSSVAVSC----- 41  
 DB 188 QNKKVEFKIDIVLAFOKASSIYKKEGEVFSFPLAFT-VEKLTGSGELMWQARAS 246  
 42 -----KASANNFRS-----YAFTWBQAPQGLQW----- 66  
 QY 247 SSKSMITFDLKNKEVSVKRVTDPRKLMGKKLPLHLTPQALPOYAGS--GNLTALAEAK 304  
 67 -----MGGIIPNF-----GAPYNAQNFQORVITISADS 94  
 DB 305 TGKHLQEVNLVVMRATOLQ--KNLT-----CEVWG-----PTSPKLM-L 340  
 95 TTYVYMETLSLTFEDTAFYYCGRGILTYGSGYYYLQHMGGTLVTVSASPTSPKYPPL 154  
 QY 341 SLKLENKAKYSKR-----EKPVWVNLPEAGM----- 367  
 DB 155 SLCTQPDGNVNIACLVGGFPPOEPLSVTWSESGGVTAARNPPSODASGDLTYTSSQLT 214  
 368 ---WQCLSDSGQVLLLESNIK--VLPTWSTVPCCPAPPKSCDKHTTCPELLG--PSV 419  
 DB 215 LPAQOCLAGKS-----VTCHVGHYTNPSODVTPPCVPSTPTPTSPPTPTSPSCCHRL 270  
 420 FLPPPKPDTLMISRTPEVTCVVVDVSHEDPEVKFNNVYDVEVHNAKTKRREOYNSTY 479  
 271 SLHPRALSD-LLLSEANLTCTLTGL--RDASGVTFMTWTPSSGK--SAVQGPDRDLGGCY 326  
 QY 480 RVSVVLTVLHDMVLANGKEYKCKVSNKALPAPIETKISAKQPREPOVYITTPSRDELTK 539  
 327 SVSSVLPGCAEPNMHGKFTCTAAYPESKPTLTATLSNS--GNTRPREVHLPPRSEELAL 385  
 QY 540 NQ--SLTCLVKGFPSPDIADAVESNGQ--PENNYKTPPVLD--SDG--SFLYSLKLTVDK 593  
 DB 386 NELVTLTCLARGFPPKDVLVNMLQSGDELPREKILTMASROBPQGTTFPNTSILNVA 445  
 QY 594 SRMOQGVFSCSVNHEALHNHYTKSLSPG-----YLDDETQ 632  
 446 EDWKKGDTFSCVGHGHALPLAFTQKTIDRLAKRPTHVNVVMAEVDGTC 495

Db 446 EDWKKGDTFSCVGHGHALPLAFTQKTIDRLAKRPTHVNVVMAEVDGTC 495

RESULT 64  
 Q91WR1 PRELIMINARY; PRT; 488 AA.

AC Q91WR1; 19, Created)  
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)  
 DT 01-DEC-2001 (Tremblrel. 19, last sequence update)  
 DE Hypothetical protein.  
 GN IGH-VJ558 OR A1893585.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OC NCBI\_TaxID=10090;  
 RP [1]  
 RA SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC013539; AAI13539.1; -  
 DR MGD; MGI:96486; Igh-VJ558.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; Ig; 4.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.

KM Hypothetical protein.  
 SQ SEQUENCE 488 AA; 52964 MW; F12068460B400B9D CRC64;

Query Match 8.4%; Score 288; DB 11; Length 488;  
 Best Local Similarity 23.8%; Pred. No. 1.3e-13;  
 Matches 128; Conservative 72; Mismatches 179; Indels 158; Gaps 29;

QY 147 PGSSPSVQCRSGKNIGGKTLVSQLELDGSGTWTCTVLQNKRYEF 195  
 33 PGASVLSKCA-SGYTTIDVYVMVWVQSHKSLLEWIDIPNYGCT-----SYNQK---F 83  
 DB 196 KIDIVLAFOKASSIYKKEGEVFSFPLAFTVEKLTGSGELMWQARASSSKWITFD 255  
 84 K-GKATLTVDKSSSIAYWQ-----LNNLTSDSAVYVYCARGPVYYSYPSYD 128  
 QY 256 LKN--KEVSVKRYTQDPKLMGKKLPLH-LTLQALP-----QYAGSNLTAL 301  
 129 RGYWGGTLVTVSAEP---AREPTIYPLTFQALSSDVIITGCLIHDFPSTGM--- 180  
 QY 302 EAKTKLHQEVNLVVMRATOLQKNLTCEVWGFPSPKMLSLKLENKAKYSKKEKPVWL 361  
 181 -----NVT--WG-----KSGKIDITTV 194  
 QY 362 N-PEAGMWQCLSDSGQVLLLESNIKVLPTWSTVPCCPAPPKSCDKHT-----TC 410  
 195 NFPRA-----LASGRYTWSQL-TLPA-----VECEGSGVSCVGHSDNPQELNVNC 243  
 QY 411 PELIGSGSVLPFPKPK-----DTLMISRTPEVTCVVVDVSHEDPE-VKFNWY 457  
 244 FGICSPPTT---PPPSQPSLSLORPALDILLGSDASTICTLNLG--RDPEGAVPTW- 297  
 QY 458 VDGVEVHNAKTKRREOYNST---YRVSVTVLTVLHDMVLANGKEYKCKVSNKALPAPIETK 514  
 298 ----EESTGDAVQKAVQNSCCYSVSVLPGCAEWMNSGASFCTVTHPESDT-LTGT 352  
 QY 515 ISRAKQPREPOVYITTPSRDELTKNQ-VSLTCLVKGFPSPDIADAVESNGQ--PENNY 570  
 DB 353 IAKVTVNTFPQVHLPPRSEELALNELVSLTCLVAFNKFVLRVRLHNGNELSPSYL 412  
 QY 571 KTRPVLDSDG--SFLYSLKLTVDKSRMOQGVFSCSVNHEALHNHYTKSLSPG 625  
 413 VFEPLEKPEGGATTVLTVSVLRVSAEIMKGGDQYSQVGHGHALPMFTQKTIDRLSG 469

Db 413 VFEPLEKPEGGATTVLTVSVLRVSAEIMKGGDQYSQVGHGHALPMFTQKTIDRLSG 469

RESULT 65  
 Q9UP60 PRELIMINARY; PRT; 384 AA.

AC Q9UP60; 13, Created)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DT 01-MAY-2000 (Tremblrel. 13, last sequence update)  
 DE SMC73 protein.  
 GN Homo sapiens (Human).  
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxID=9606;  
 RP [1]  
 RA Zheng S., Cao J., Cao W., Cai X., Geng L.;  
 RT "Identification and characterization of SMC73, a gene which is down-  
 regulated in colorectal cancer";  
 RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AF067420; AAC19365.1; -  
 DR HSSP; P01825; 7FAB.  
 DR InterPro; IPR007110; IG\_1like.  
 DR InterPro; IPR003597; IG\_C1.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR Pfam; PF00047; Ig; 3.  
 DR SMART; SM00407; IGC1; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 3.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 SQ SEQUENCE 384 AA; 40947 MW; BA7AD3CA5A9DD48 CRC64;

Query Match 8.3%; Score 282.5; DB 4; Length 384;  
Best Local Similarity 26.5%; Pred. No. 2.5e-13;  
Matches 92; Conservative 46; Mismatches 136; Indels 73; Gaps 12;

QY 305 TGLHDEVLVVRATQ--LQKNLCEVWGPTSPKMLSLKLENKEAKYSKEKPVWLN 362  
DB 91 SGLVLTSSQGLTPAQCCLAGKSVTCHVGHYTNP----- 124

QY 363 PEAGMOCCLISDSGVLLSNNIKVLPFTWSTPVPCCPAPPEKSCDKHTCELLAGSEVLP 422  
DB 125 -----SQDVTVPCEVPTPTSPPTSPPTSPSCCH-----PRLSLH 161

QY 423 PPKKDTLMISRPETVCVVVDVSHEDPEVKFNWYDGVVHNAKTKPREEOYNSTRYV 482  
DB 162 RPLLED-LILGSEANLITLTGL-RDAGVTFTWTPSSGK--SAGVGPREDICGYSYS 217

QY 483 SVLTVHODMLNGEKYCKVSNKALPAPIEKTISKAKGPREFQVYTLPPSRDELTKNO- 541  
DB 218 SVLPGCAEPNNHCKTCTTAAYPESEKTPLATLSKS-GMTFREVLHLPSPSEELANL 276

QY 542 VSLTCLVKGFPSPDIIVEMESNGQ--PENNYKTPPVLD-SDG--SFPLYSKLTVDKSRW 596  
DB 277 VTLTCLARGSPKDVIVRWLQSGQELPREKYLTMASRGEPGQTTFAVTSILRVAAEDW 336

QY 597 QQGNVFSQVMEHALNHYTKSLSPG-----LQDENC 632  
DB 337 KKQDTSCVWGHEALPLAFTQKTIIDRLAGKPHVNVSVMAEDGTC 383

RESULT 66  
080217 PRELIMINARY; PRT; 487 AA.

AC 080217;  
DT 01-JUN-2003 (TREMBLrel. 24, Created)  
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=FVB/N; TISSUE=Colon;  
RA Strausberg R.;  
RL Submitted (MAR-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC049143; AA049143.1; -  
DR InterPro; IPR003599; IG.  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003597; IG-cl.  
DR InterPro; IPR003006; IG\_MHC.  
DR Pfam; PF00047; IG\_3.  
DR SMART; SM00409; IG\_3.  
DR SMART; SM00407; IG1; 3.  
DR SMART; SM00406; IG1; 1.  
DR PROSITE; PS50835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
KW Hypothetical protein.  
SQ SEQUENCE 487 AA; 53019 MW; 31F2C893900A4D80 CRC64;

Query Match 8.3%; Score 282.5; DB 11; Length 487;  
Best Local Similarity 22.9%; Pred. No. 3.5e-13;  
Matches 131; Conservative 74; Mismatches 197; Indels 169; Gaps 29;

QY 121 LVFGLANSDBTHLLOGSLTTLTLES-----PPGSSPSVQCRSPRGNIQGGKTLVSQDL 174  
DB 1 MYGLNCVFLVFLTKGVQSEVKLEAGGGLVQPGSKUSCAA-----SGFTF----- 48

QY 175 ELQDSGTWCTVLQN-QKAYEPKIDIVLAFQKASSIVYKKEGQVEFSPFLAFVTEKLT 233  
DB 49 -----SNVMMWNVQSPKGLWEVAEIRLRNNNATHTYAESVKGR-----FTISRSD 95

QY 234 GSGELMWQ-----AERASSSKSMTFEDLKNKESVRYVTDPKLQMGKKLP 279  
DB 96 SKGSYVLQMNRIARDGTGYICTRRGIDPPMY-FDWAGATTV-TVSEPP---AREPT 149

QY 280 LH-LTLPOALP-----QVAGSGLTLALAKTGKLGQEVNLVVRATQLOKIL 327  
DB 150 IYPLTFPQALSDPVIIGCLIHDPFGSGTM-----NVT 182

QY 328 CEVWPTSPKMLSLKLENKEAKYSKEKPVWLN-PEAGMOCCLISDSGVLLSNNIKV 386  
DB 183 ---WG-----KSGKDIITVFPFA-----LASGRYTMSSQL-T 212

QY 387 LPTWSTPVPCCPAPPEKSCDKTH-----TCPELLGSEVFLFPPEK----- 427  
DB 213 LPA---VECPGSEVSKSVQHDSPVQELVNCGICSPPT---PPPSQPSLSIOR 265

QY 428 ---DTLMTISRPETVCVVVDVSHEDPE-VKFNWYDGVVHNAKTKPREEOYNST---YR 480  
DB 266 PALLEDLLIGSDASITCTINGL-RDEGAVFTW-----EPSTGDAVQKAVQNSCGYS 318

QY 481 VSVTLVHODMLNGEKYCKVSNKALPAPIEKTISKAKGPREFQVYTLPPSRDELTKN 540  
DB 319 VSSVLPGCAERNWNSGASFKCTVTHPSDT-LTGITAKYVNTFPQVHLLPPSEELANL 377

QY 541 Q-VSLTCLVKGFPSPDIIVEMESNGQ--PENNYKTPPVLDSDG--SFPLYSKLTVDKS 594  
DB 378 ELVSLTCLVRAPNPKVIVRWLHGNBELSPSYLVFEPKKEGQATTTLVSLRVSAE 437

QY 595 RMQGNVFSQVMEHALNHYTKSLSPG 625  
DB 438 IKWQGDQYSQVWGHEALPNNFTQKTIIDRLSG 468

RESULT 67  
08WY24 PRELIMINARY; PRT; 497 AA.

AC 08WY24;  
DT 01-MAR-2002 (TREMBLrel. 20, Created)  
DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)  
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE SNCG6 protein.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Zheng S., Shao X., Cao J., Geng L., Fang Y., Dong Q.;  
RT "Identification and characterization of SNCG6, a Ig-like gene which is  
RT down-regulated in colorectal cancer.";  
RL Submitted (JUN-2000) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF283666; AAL36987.1; -  
DR InterPro; IPR007110; IG-like.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; IG\_4.  
DR SMART; SM00406; IG1; 1.  
DR SMART; SM00405; IG1; 1.  
DR PROSITE; PS50835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 1.  
KW  
SQ SEQUENCE 497 AA; 53665 MW; F24D08DFA5A63E5 CRC64;

Query Match 8.3%; Score 282.5; DB 4; Length 497;  
Best Local Similarity 25.3%; Pred. No. 3.6e-13;  
Matches 117; Conservative 60; Mismatches 205; Indels 81; Gaps 20;

QY 223 VEKLGGSEGLMWQARASSSKSMTFEDLKNKESVRYV---RVT---ODPKLQMGKKLP 283  
DB 56 VQAPRGQGLEW-----MGWMPQGTGNTFPAQKFGKRLPSRPTSINTAYMVSLS 106

QY 284 LPQALPOVAGSGL-----TLAEAKTGKLGQEVNLVVRATQLOKN- 325  
DB 107 TEDSAIYFCARNTLGGRGFGVNWDFPMWGHGLTVVSSASPTNPVKVFLSLCSTQPDGNY 166

```

Oy 326 -LTCEVMG--PTSPKLMSTLKENKAYSKREKPVWVLNPEAGMOCCLSDSGVLLS 382
Db 167 VIACLVGFFPOE---LSVTWSESGGVTAARNP-----PSQASGLTYTSSQTLIPA 218
Oy 383 -----NIK--VLPTWSTVPCCAPAPKSCDKHTHTPELLG--PSVELFPKPK 426
Db 219 TOCLAGSVTCHVGHYTNPSQDVTVPVCPSTPTPTSPSTPTPTSPSCCHRLSLHRL 278
Oy 427 KDTLMISTPEVTCVVDVSHEDPEVKWVVDGVENHATKXREOYNSTYRVSVLT 486
Db 279 ED-LILGSEANLTCTLTGL-RDASGVPTFTWTPSSGK--SAVQGPERRDLCCGYSVSVLP 334
Oy 487 VLHODWLVGKFKYKCVSNKALPAPIEKTISKAKQPREPOVYLTTPSRDELTKNQ-VSLT 545
Db 335 GCAEPMNHGKFTCTTAAPESKTYLTATLPKS-GNTPPEVHLLPPSEELALNELVLT 393
Oy 546 CLVKGFPSPDIIVEMESNGQ--PENNYKTPPVLID-SDG--SFLYSKLTVDKSRMOQN 600
Db 394 CLAGFSPKQVLMVWLGSGQLPREKYLTMASROEPSCGTTFAVTSLRVAADMKKGD 453
Oy 601 VFSCSVNHEALHNYTKSLSLSPG-----LQLEETC 632
Db 454 TFCMVGHGHALPLAFTOKTIDRLAGKPTHVNVSVVMAEVDGTC 496

RESULT 68
ID Q29028 PRELIMINARY; PRT; 99 AA.
AC Q29028;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Cda, allele 2 (Fragment).
OS Sus scrofa (pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
ON NCBI_Taxid=9623;
RX STRAIN=FROM N.A.
RA MEDLINE=9332916; PubMed=8335933;
RT "Extensive allelic polymorphism in an exposed region of the
RT mlature. . . . .";
RL J. Immunol. 151:1365-1370(1993).
DR EMBL; X65630; CAA46584.1; -.
DR PIR; I47132; S21462.
DR HSRP; P01730; ICDY.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00406; Ig; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
FT NON_TER 1
FT TER 99
SQ SEQUENCE 99 AA; 11390 MW; C229B5BEA228318F CRC64;

Query Match 8.2%; Score 280.5; DB 6; Length 99;
Best Local Similarity 56.1%; Pred. No. 5,1e-14;
Matches 55; Conservative 19; Mismatches 23; Indels 1; Gaps 1;
Oy 32 KGGTVELTCTASOKSIQPHWKNNOIKIIGNOSFLTKGP-SKLNDRADSRSLMDQG 90
Db 1 KAGGLAELPCHSSQKKNLPFSWKNSDQIKILRSHRNLMHKASVTELSRLDSKKMMWDHG 60
Oy 91 NFPLIINKLKIEDSDTYICEVEDQKEXEQLVPLGLTAN 128
Db 61 SFLIINKLEVTDSGIYICEVEDKIEVQLVPLRLTAS 98

```

```

ID Q8VCX4 PRELIMINARY; PRT; 489 AA.
AC Q8VCX4;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
GN IGH-VJ558 OR A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
ON NCBI_Taxid=10090;
RX STRAUSBERG R.;
RA Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RL EMBL; BC018322; AAH18322.1; -.
DR MGD; MGI:96486; IGH-VJ558.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003596; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00406; Ig; 1.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 489 AA; 53208 MW; CC85B1194DAFEF2C CRC64;

Query Match 8.2%; Score 278.5; DB 11; Length 489;
Best Local Similarity 23.9%; Pred. No. 7.2e-13;
Matches 137; Conservative 71; Mismatches 191; Indels 175; Gaps 32;
Oy 118 VOLIVFGLTN--NSDTHLLO-GOSLTTLTSPSSPSVOCRRPKGNIGKGLTYSQ 174
Db 6 VFPLFLSVTAGVHSKVQLQSGAELV-----KFGASVKSCKA-----SGYTFSDYFI 53
Oy 175 E--LQSGT-----W-----TCVTLQNKQKVEPKRIDIVLAFKASSIVYKKGGEQVEFS 222
Db 54 HMKQRSGGLEWIGWNPSSGSIKFNEK--FK-DATLTADKSTTV----- 99
Oy 223 FPLAFTVEKLTGSGELWMAER-----ASSSKWTFIDKKNREVSVK-RVTDPKLQMG 275
Db 100 -----MDLSRLTSDSAVFPARHEDKNGVSLAMFWYMQGLVTVYSAPAPAEPTI--- 152
Oy 276 KKLPLHLTLPQALP-----QYAGSGNTLALKATGKLHDEVNLVNRATQLQK 324
Db 153 -----YPLTFPQALSDPVIIGCLIHDFPFGTM----- 181
Oy 325 NLTCFVWVGPTSPKLMSTLKENKAYSKREKPVWVLN-PEAGMOCCLSDSGVLLS 383
Db 182 NVT--WG-----KSGKDIITVNFPPA-----LASGRRYTMSQ 212
Oy 384 IKVLPWTSTVPCCAPAPKSCDKTH-----TCPELLGSPVFLFPKPK----- 427
Db 213 L-ILPA-----VECPREGESVVCVQHSNRPVELANNCFGICSPPT---PPPSQPSLS 264
Oy 428 -----DTLMIISRTPEVTCVVDVSHEDPE-VKFNMYVDGVENHATKXREOYNST- 478
Db 265 LQRPALLEDLLGSDASITCTLNGI--RDPEGAFTW-----EPTSGDAVQKAVQNSCG 317
Oy 479 -YRVSVGLTVLHODWLVGKFKYKCVSNKALPAPIEKTISKAKQPREPOVYLTTPSRDEL 537
Db 318 CYSVSVSVLPCCARWMSGSAFKCTVTHPESDT-LTGITAKVTYNTPEPQVHLLPPESEL 376
Oy 538 TKNQ-VSLTCLVGVFPSPDIIVEMESNGQ--PENNYKTPPVLIDSG--SFLYSKLTIV 591
Db 377 ALNELVSLTLVAAPKQVLMVWLGNELSSESLVPEPLKPEBEGATTIVTVLVRV 436
Oy 592 DKSRMOQGNVSCSVNHEALHNYTKSLSLSPG 625
Db 437 SAEIWKQGDQYSCMVGHGHALPMNFTQKTIDRLSG 470

```

RESULT 69  
Q8VCX4

RESULT 70  
091WT1 PRELIMINARY; PRT; 481 AA.

AC 091WT1;  
DT 01-DEC-2001 (TREMBLrel. 19, Created)  
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Hypochemical protein.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Colon;  
RA Strauberg R.;  
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC013490; AH13490.1; -  
DR InterPro; IPR007110; IG\_1like.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; IG; 4.  
DR SMART; SM00406; IG; 1.  
DR PROSITE; PS50835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
KM Hypochemical protein.  
SQ SEQUENCE 481 AA; 52105 MW; 97DF68D159463F65 CRC64;

Query Match 8.1%; Score 276; DB 11; Length 481;  
Best Local Similarity 21.9%; Pred. No. 1.1e-12;  
Matches 140; Conservative 67; Mismatches 207; Indels 226; Gaps 25;

15 QLLALPRAATGANKVGLKKDYLTLCTAS--OKSIQPFMKNSNOIKILGNG----- 66  
20 QVOLLQSGPE---LVKPGASVAKISCKASGYFTSYTHW-----VKORPGGLVWIGW 69  
67 SFLTKGSKLNDRAISRSLW---DQGNFPLIKNLKIEDSDTYICEVEDQKEVOLVLF 123  
70 IYPGDNTKTKNEKFKKTTITADKSSSTAYMFLSLTSEDSANYFC----- 115  
124 GLTANSDTHLLQGSQTLTLTLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWT 183  
116 -TRGGGWAFDYMGQGTTLTVSSSEPARPT----- 143  
184 CTYLQKQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSPLAFYVEKLTGSGELMWQAE 243  
144 -----IYPLTFPQALSSDPYIIGCLHNDYFP-----SGTNNVTW--- 177  
244 RASSSKSWITFDLKNKEVSVKRVTOPDKLQMGKKLPLHLTLPOALPOYAGSGNLT-AL 302  
178 -GKSGK-----DITVNFPPALASGGR-----YTMSSQLTLPAYE 211  
303 AKTGKLGHEVNLVVMRATQLOKMLTCEWGPSTPKMLSLKLENKAKVSKREKPVAVLN 362  
212 CPEG-----ESVVCSSVQHDNSNPQELN 233  
363 PRAGMMOCLSDSGQVLESNIKVLPTWSTPVPCP-----APEKSCDKHTTCBELLG 415  
234 -----VNCFGICSPPTTPPPSCQ----- 252  
416 GPSVLPFPKPKDMLMISRTPEVTVGVVDVSHEDPE-VKENMYVVDGVEVNAKTKPREBO 474  
253 -PFLSLORPALLE-LIIGSDASITCTINGL--RDPEAVFTW-----EPSTGDVAQKKA 303  
475 YNST--YRVSVYLVLDHMDLNGKEYCKVSNKALPAPIEKTISRAGQPREPOVYTL 511  
304 VQNSCGYSVSVLPGCAERWNSGASPKCTVTHPESDT-LTGTLAKTVVTPPQVHL 362  
532 PSDELTKQO-VSLTCLVKGFPSPDIANVESNGQ--PENNYKTPPVLDSDG--SFL 585  
363 PRBEELALNELVSLTCLVRAFNPKEVLYRWLHGNBELSPESYLVFEPLKEPREGATTV 422  
586 YSKLTVDSKRWQGNVSCSVMEALHNHYTOKSLSPG 625

Db 423 TSVLRVSAEIMWKQDQYSCMWGHEALPNMFTGKTIDRLSG 462

RESULT 71  
08K172 PRELIMINARY; PRT; 482 AA.

AC 08K172;  
DT 01-OCT-2002 (TREMBLrel. 22, Created)  
DT 01-OCT-2002 (TREMBLrel. 22, Last sequence update)  
DE 01-OCT-2003 (TREMBLrel. 25, Last annotation update)  
DE Similar to expressed sequence A1893585.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC TISSUE=Breast tumor;  
RA Strauberg R.;  
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; BC028249; AH28249.1; -  
DR PIR; F33932; F33932.  
DR PIR; PH1105; PH1105.  
DR PIR; PH1108; PH1108.  
DR PIR; PH1114; PH1114.  
DR PIR; PH1118; PH1118.  
DR PIR; PH1119; PH1119.  
DR PIR; PH1125; PH1125.  
DR PIR; PH1126; PH1126.  
DR PIR; PH1128; PH1128.  
DR PIR; PH1131; PH1131.  
DR PIR; PH1134; PH1134.  
DR PIR; PH1139; PH1139.  
DR PIR; PH1142; PH1142.  
DR PIR; PH1149; PH1149.  
DR PIR; PH1150; PH1150.  
DR PIR; PH1151; PH1151.  
DR PIR; PH1152; PH1152.  
DR PIR; PH1153; PH1153.  
DR InterPro; IPR007110; IG-1like.  
DR InterPro; IPR003597; IG\_c1.  
DR InterPro; IPR003006; IG\_MHC.  
DR InterPro; IPR003596; IG\_v.  
DR Pfam; PF00047; IG; 4.  
DR SMART; SM00407; IGc1; 3.  
DR SMART; SM00406; IG; 1.  
DR PROSITE; PS50835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 2.  
SQ SEQUENCE 482 AA; 52121 MW; A06F083E771D084 CRC64;

Query Match 8.1%; Score 276; DB 11; Length 482;  
Best Local Similarity 22.0%; Pred. No. 1.1e-12;  
Matches 137; Conservative 68; Mismatches 203; Indels 214; Gaps 24;

30 LGKGDYVELCTAS--OKSIQPFM---KNSNOIKILGNSQFLTKGSKLNDRAISR 84  
30 LVKPGASVAKISCKASGYFTSYTHWVVKORPGKLEWIGRIDP--NSGKTKNEKFKSA 87  
85 SLW---DQGNFPLIKNLKIEDSDTYICEVEDQKEVOLVGLTANSDTHLLQGSQTL 141  
88 TLTVKPSSTAYMQLSLTSEDSANYFCRE-----GDVYAMDYMGGTSTV 133  
142 TLESPGSSPSVQCRSPRGKNIQGGKTLVSQLELDSDGTWTCTYLQKQKVEFKIDIV 201  
134 TVSSSEPARPT-----IYV 147  
202 LAFQKASSIVYKKEGEQVEFSPLAFYVEKLTGSGELMWQARASSSKSWITFDLKNKEV 261  
148 LTFPQALSSDPYIIGCLHNDYFP-----SGTNNVTW-----GKSGK----- 183  
262 SVKRVTOPDKLQMGKKLPLHLTLPOALPOYAGSGNLT-ALBAKTKGLHGEVNLVVMRAT 320



```

Db 184 DTTNFPALASGR-----YTWSSQLTLPVACEEG-----216
Qy 321 OLQKMLTCEVWGPTSPKMLSLKLENKEAKYSKREKPYVWVLPAGMMQCLLSDSGVYL 380
Db 217 -----ESVKCSYQHNSNPQELN-----234
Qy 381 ESNIKVLEWSTWPCP-----APEBKSCDKHTHCEBLLGSPSVLFPKPKDTLMS 433
Db 235 -----VNCPCIGSPPTTPPSCQ-----PSLSQRLALBD-LTLG 269
Qy 434 RTEPVTGVVVDVSHEDPE-VKFNMYVDGVVHNAKTKREQVNST--YRVSVLTVLH 489
Db 270 SDASITLTNLNL--RDEGAVFTW-----EPSTGKDAVQKAVONSCGCVSVSLDPCA 322
Qy 490 QDMNGKEKCKVSNKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKNQ-VSLTCLV 548
Db 323 ERNMSGASFCTVTHPESDT-LTGITAKVYVNTTPPOVHLPPSEELALNELVSLTCLV 381
Qy 549 KGFYPSDIAVWESNGQ---PENNYKTPPVLDSDG--SFLYSKLTVDKSRMOQGNVFS 603
Db 382 RAFNPKVLYVWMLHGNBELSPESYLVPFLKEPEGATTYLVTSVLRVSAEIMWQSDQYS 441
Qy 604 CSVMHEALHNHYTQKSLSLSPG 625
Db 442 CMVGHEALPMNFTQKTIDRLSG 463

```

## RESULT 72

```

Q99KA4 PRELIMINARY; PRT; 487 AA.
ID Q99KA4
AC Q99KA4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC004786; AAH04786.1; -
DR HSSP; P01810; 2PBJ.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 4.
DR SMART; SMO0406; Ig; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ
SEQUENCE 487 AA; 52554 MW; 7DC8B96DB333077B CRC64;

```

Query Match 8.0%; Score 274.5; DB 11; Length 487;  
 Best Local Similarity 21.7%; Pred. No. 1,4e-12;  
 Matches 141; Conservative 81; Mismatches 221; Indels 207; Gaps 27;

```

Qy 1 MNRGVPRHLLLV---QLALLPATQGNKVLGKSGDTVELTCTAS--QKSIQTFM-- 53
Db 1 MNFPLSLFLVLVLYKQVCEVQVLESGLV---KPGSLKLSCLNAGGFTFSSVAMSWVR 57
Qy 54 ---KSNQIKILNQGSFLTKGPKLNDRAISRSLMDQGNFPLIKNLIKIEDSDTYIC 109
Db 58 QTPKRLKEMVATISDGSY--TYPDNVKGRFTISRDN-AKNILYLQMSHLKSEDTAYYC 115
Qy 110 EVEQKKEVQLLVGLTANSDTHLLQGSGLTLTLESFPGSSPSV-QCRSPRGKTIQAGKT 168
Db 116 ---ARMGSPYGVGYSRFD---YWGQGTITVSSSARAKPTIPLTPRA----- 159
Qy 169 LSVSLELQDSGVTCTVLQNKQKVEFKIDIVLAFQKASSIVYKKEGEQVEFSFPLAFT 228

```

```

Db 160 -----LSDDPVII-----GCLIHDFP-----176
Qy 229 VEKLTGSGELMWQABRASSKSWITPDLKXKEYSVKRVTPQPKLQMGKULPLHLTPQAL 288
Db 177 -----SGTMNVYWGSKSDITTT-----VNFPAL 200
Qy 289 PQ---YAGSNLTL-ALEAKTGKLEHVNLVWRA-TOLQKMLTCEVWGPTSPKMLSLKL 344
Db 201 ASGGGYMSSQLTLPAVEEC-----219
Qy 345 ENKEAKYSKREKPYVWVLPAGMMQCLLSDSGVLLSNIKVLPTWSTWPCPAPBPKSC 404
Db 220 -----PEGESVCKSVQSDNAVDELVKCS---GPPPCP-PCPPSC 257
Qy 405 DKHTHCEBLLGSPSVLFPKPKDTLMSITPVTGVVVDVSHEDPEVKRWVVDGEVH 464
Db 258 H-----PSLSQRLALBD-LTLGSDASLTCLNLNRPBGAV-FTW-----EPS 299
Qy 465 NAKTKPREQVNST--YRVSVLTVLHODMNGKEKCKVSNKALPAPIEKTISKAKQ 521
Db 300 TGDKDAVQKAVQNSCGCVSVSLPGCAERKMSGASFCTVTHPESDT-LTGITAKTVN 358
Qy 522 PREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPYSDIAVWESNGQ---PENNYKTPPV 577
Db 359 TFPPOVHLPPSEELALNELVSLTCLVRAFNPKEVLYVWMLHGNBELSPESYLVPFLKE 418
Qy 578 DSDG--SFLYSKLTVDKSRMOQGNVFSVMHEALHNHYTQKSLSLSPG 625
Db 419 PGGATTYLVTSVLRVSAETWQSDQYSCMVGHEALPMNFTQKTIDRLSG 468

```

## RESULT 73

```

Q8VEAO PRELIMINARY; PRT; 484 AA.
ID Q8VEAO
AC Q8VEAO;
DT 01-MAR-2002 (Tremblrel. 20, Created)
DT 01-MAR-2002 (Tremblrel. 20, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
GN IGH-VJ558 OR A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC019425; AAH19425.1; -
DR MGD; MGI:96486; Igh-VJ558.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 4.
DR SMART; SMO0406; Ig; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KW Hypothetical protein.
SQ
SEQUENCE 484 AA; 52859 MW; F2522DFF5ED6288A6 CRC64;

```

Query Match 8.0%; Score 274; DB 11; Length 484;  
 Best Local Similarity 24.8%; Pred. No. 1,6e-12;  
 Matches 102; Conservative 58; Mismatches 153; Indels 98; Gaps 20;

```

Qy 289 PoyAGS--GNLTALBAKTGLHQAENVLVWRA-----TOLQKMLTCEVWG----- 332
Db 79 PEYSASVQGRFTISRNSQNIILYLNW--TLRAEDATYYCAARAYSNYYEDNMNGQSTTL 136
Qy 333 -----PTSPMLSLKLENKAVY-----SKREKPYVWVLA-PEAG 366
Db 137 VSSEPAAREPTIPLTPQALSSDPVILIGCLIHDFPSGTMNVYWGSKGKIDITVNPFA- 195
Qy 367 MMQCLLSDSGVLLSNIKVLPTWSTWPCPAPBPKSCKTH-----TCPBLLG 416

```

```

Db 196 -----LASGRYMSQL-TLPA-----VECPBESVKCSVQHDNSNVQELANNCPGICSP 245
QY 417 PSVFLFPKPK-----DTLMSRTPEVTCVVVDVSHEDPE-VKPMVYDGVAV 463
Db 246 PTT---PPPPSCQPSLQRPALDELILGSDASITCTLGL--RDEGAVFTW-----EP 295
QY 464 HNAKTPREEQNST---YRVSVLTVLHQDLNKEGYKCKVSNKALPAPIEKTISKAG 520
Db 296 STGKDAVQKKAQVNSCGCYSSVSLPGCAERNMSGASFCTVTHPESDT-LTGTLAKTV 354
QY 521 QPREPQVTLPPSRDELTKNQ-VSLTCLVKGYFSPDAVEMESNQO---PENNYKTPPV 576
Db 355 NTFPPQVHLPPSEBELINELVSLCLVRAFPKCVLVRMLHGNELSPESIVPEPLK 414
QY 577 LQSDG--SFFLYSKLTVDSKRWQGVNFCGVNHEALHNHYTQKSLSLSPG 625
Db 415 EPQEGATYLVTSVLRVSAEIMKQDQYSCMVGHEALPMNFTQKTRLSG 465

```

## RESULT 74

```

Q91WT3 PRELIMINARY; PRT; 481 AA.
ID 091WT3
AC 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Hypothetical protein.
GN MGI-VJ558 OR A1893585.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP TISSUE=Colon;
RA Strausberg R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL, BC013488; AAL13488.1; -.
DR MGI; MGI:96486; Igh-VJ558.
DR InterPro; IPR007110; I9-1like.
DR InterPro; IPR003006; I9_MHC.
DR InterPro; IPR003596; I9_V.
DR Pfam; PF00047; I9_4.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS00835; I9_LIKE; 4.
DR PROSITE; PS00290; I9_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 481 AA; 52022 MW; 4EB5C253038B718 CRC64;

```

```

Query Match 8.0%; Score 272; DB 11; Length 481;
Best Local Similarity 22.6%; Pred. No. 2.2e-12;
Matches 123; Conservative 69; Mismatches 168; Indels 184; Gaps 26;

QY 182 WTCTV-----ONOKVFEKIDIVLAFQKASIVYKKEGQVESFPLA---FT- 228
Db 3 WSCIILFLVATANGVNSQVQL-----QPGAEIV--RPGAIVKLSCKTSSTTFID 50
QY 229 -----VEKLTSGLSLWQA-----ERASSSKSWITFDLKNKEVS 262
Db 51 YMMNWVKQRPQGLQLEWIGALDPDSYTSYNGKFKGTTLVDTSSSAYMLLSLTSEDSA 110
QY 263 YKAVTQDPKQMG-----KKLPLHLTLPLALP-----QYAS 294
Db 111 VYFCAGAPRDSQSYWGGTTLVSSBPAREPTIYLPFPQALSSDPVLIIGCLIHDPFS 170
QY 295 GNLTLLEAKTGKLGHEVNLVWRATQLOKNTLCEWNGPFSPLMLSLKLENKAIVSR 354
Db 171 GTM-----NVT---WG-----KS 180
QY 355 EKEVWVILN-PEAGMOCQLSDSQVLLESNIKVLPTWSTPVPAPAPKSCDKTH----- 408
Db 181 GKQITTVNFPAP-----LASGRYMSQL-TLPA-----VECPBESVKCSVQHDNSNPV 229

```

```

QY 409 -----TCBELIGSPVFLFPKPK-----DTLMSRTPEVTCVVVDVSHEDPE 451
Db 230 QELNVCPCICSPPTT---PPPPSCQPSLQRPALDELILGSDASITCTLGL--RDE 284
QY 452 -VKPMVYDGVAVHNAKTPREEQNST---YRVSVLTVLHQDLNKEGYKCKVSNKAL 507
Db 285 GAVFTW-----EPSTGDAVQKKAQVNSCGCYSSVSLPGCAERNMSGASFCTVTHPE 339
QY 508 PAPIEKTISKAGQPREPQVTLPPSRDELTKNQ-VSLTCLVKGYFSPDAVEMESNQO- 565
Db 340 DT-LTGTLAKTVNTFPQVHLPPSEBELINELVSLCLVRAFPKCVLVRMLHGN 398
QY 566 ---PENNYKTPPVLDSDG--SFFLYSKLTVDSKRWQGVNFCGVNHEALHNHYTQKS 621
Db 399 LSPESIVLPEPLKEPBGATTVLVTSVLRVSAEIMKQDQYSCMVGHEALPMNFTQK 458
QY 622 LSPG 625
Db 459 RLSG 462

```

## RESULT 75

```

Q91X92 PRELIMINARY; PRT; 482 AA.
ID 091X92
AC 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE Hypothetical protein.
GN PIR; F33932; F33932.
DR InterPro; IPR007110; I9-1like.
DR InterPro; IPR003006; I9_MHC.
DR InterPro; IPR003596; I9_V.
DR Pfam; PF00047; I9_4.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS00835; I9_LIKE; 4.
DR PROSITE; PS00290; I9_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 482 AA; 51865 MW; 312E01F9C1BC7F3C CRC64;

```

```

Query Match 8.0%; Score 272; DB 11; Length 482;
Best Local Similarity 22.1%; Pred. No. 2.2e-12;
Matches 143; Conservative 69; Mismatches 218; Indels 216; Gaps 26;

QY 10 LLLVQLALPRAATGKRVYLGKG-----DTVELCTAS--QKSGIQFMKNSNQK 60
Db 4 LALLCLVTFPSGAL--SQVLKESGPDVVAQSISITCTVSGFALTSAISWAPQPEK 62
QY 61 ILNGQSPFLTKGPKSLNDRADRSRL-WDOGNFPLIIRNLKIEDSDT--YICEVEDQKE 117
Db 63 GLEMLGVMTGQVTNYSALSKSRSLISKDQSKSQVFLKQNSLQNTDTRAYC-AQDSNVE 121
QY 118 VQLLVFGLTANSQTHLQGSITLTLBSPGSSPSVQCRSPKGNIOGKTLVSQLEIQ 177
Db 122 GAMDYW-----GGQTSVTSSEBPAREPT----- 144
QY 178 DSGTWTCTVQLNQKVEFEKIDIVLAFQKASIVYKKEGQVESFPLAFTVEKLTSGE 237
Db 145 -----IYLPFPQALSSDPVLIIGCLIHDPF-----SGTNK 175
QY 238 LMMQAEKASSSKSWITFDLKNKEVSVKRVTDQPKLQMGKQLPLHLTLPLALQVAGSGL 297

```

```

Db      176 VTM-----GKSGK-----DITVNFPPALASGR-----YTWSSQL 206
Qy      298 TL-ALBAKTKLHOEVNLVVMRAATQLOKNTLCEWGTSPKMLSLKENKAQVSRK 356
Db      207 TLPAVECPREG-----ESVKSQVQHDN 228
Qy      357 PWWNLNDEAGWMOCLLSDSGVLLSNIKVLPTSTPVRP-----ASEPKSCDKHTH 409
Db      229 PVQELN-----VNCRGICSPPTTPPSCQ-----253
Qy      410 CPPELLGPPVFLPEPKKDTLMSRTPEVTCVVVDVSHEDPE-VKFNWYVDGVEVHNAKT 468
Db      254 -----PSLSLQRPALSD-LLIGSDASITCTLNGL--RDPEGAVFTW-----EPSTGKD 298
Qy      469 KPREQVNST---YRVSVLTVLHODWLNGEYCKYCNKALPAPIEKTSKAKGQPREP 525
Db      299 AVOKKAVONSCGCGSVSSVLPVPGCAERNNSGASFCTYTHPSDT-LTGTLAKVTNTTPP 357
Qy      526 QVYTLPPSRDELTKNQ-VSLTCLVKGFYPSDIAVEMESNQ---PENNYKTTPEVLDSG 581
Db      358 QVHLPPSEELANLVELSLTCLVRAFNPKKVALVRLHGNBELSPESYLVPEPLKPERG 417
Qy      582 --SFFLYSKLTVDKSRWQGNVFCSVMEALNHNHTYQKSLSLSPG 625
Db      418 ATYLVTSVLRVSAEIWKQGDQYSCWVGHIALPMNFTQKIDRLSG 463

RESULT 76
Q91WP5 ID Q91WP5 PRELIMINARY; PRT; 479 AA.
AC Q91WP5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Colon;
RA Straubenz R.;
RL Submitted (SEP-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC013656; AAH13656.1; -
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PSS0290; IG_MHC; 2.
KW Hypothetical protein.
SQ SEQUENCE 479 AA; 51603 MW; ECB2D087748584F CRC64;

Query Match 8.0%; Score 271.5; DB 11; Length 479;
Best Local Similarity 24.1%; Pred. No. 2.4e-12;
Matches 136; Conservative 72; Mismatches 188; Indels 169; Gaps 32;

123 FGLTANSDTHLLOQSLLTLT-ES-----PPGSSPVQCRSPRGNKIQGGKTLSSVQLEL 176
Db      3 FGLTLILVLTLKVOCEVOLVESGGGLVKFGSLKSCAA-----SGLTFSNYVM-- 53
Qy      177 QDSGTWTCVTLQN-QKVEFKRIDIVLAFOKASSIVYKKEGEQVEFSPFLAFTVEKLTGS 235
Db      54 ----SM-----VRQSPKRLW-----VAAINSNGNTYSD-----TWKGRFTISRDNK 95
Qy      236 GELMWQERASSSSKMIT-----FDLAKNEVSVKRTQDPKIQMGKLLPLH-LTL 284
Db      96 STLVLQ---WSSLSEDAFYCYVGGVFDVWAGATV-TVSSSP-----AAEPITYPLTF 147
Qy      285 PQALP-----QVAGSGLTLALFAKTKLHOEVNLVVMRAATQLOKNTLCEWGP 333

```

```

Db      148 PQALSSDPVILIGLLIHDPYPSGTM-----NVT---WG- 176
Qy      334 TSPKMLSLKENKAQVSRKPVVNLN-PEAGWMOCLLSDSGVLLSNIKVLPTWST 392
Db      177 -----KSGKIDITVNFPPA-----LASGRYTMSSQL-TLPA--- 207
Qy      393 PVPCPAPBPCKCDKTH-----TCPELLGPPVFLPEPKK-----DTL 430
Db      208 -VECPGSESVKSGVQHDNSNPVQELANVNCRGICSPPT-----PPPSQPSLSLQRPALDEL 263
Qy      431 MISTRPEVTCVVVDVSHEDPE-VKFNWYVDGVEVHNAKTKPREQVNST---YRVSVLT 486
Db      264 LLGSDASITCTLNGL--RDPEGAVFTW-----EPSTGKAVOKKAVONSCGCGSVSSVLP 316
Qy      487 VLHODWLNGEYCKYCNKALPAPIEKTSKAKGQPREQVYTLPPSRDELTKNQ-VSLT 545
Db      317 GCAERNNSGASFCTYTHPSDT-LTGTLAKVTNTTPPQVHLIPPSSEELANLVELSLT 375
Qy      546 CLVKGFYPSDIAVEMESNQ---PENNYKTTPEVLDSG--SFFLYSKLTVDKSRWQGN 600
Db      376 CLVRAFNPKKVALVRLHGNBELSPESYLVPEPLKPERGEGATTVLVSRVSAEIWKQGD 435
Qy      601 VFSGSVMEALNHNHTYQKSLSLSPG 625
Db      436 QVSCWVGHIALPMNFTQKIDRLSG 460

RESULT 77
Q90WB5 ID Q90WB5 PRELIMINARY; PRT; 482 AA.
AC Q90WB5;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor.
OS Anas platyrhynchos (Domestic duck).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Anseriformes; Anatidae; Anas.
OX NCBI_TaxID=8839;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Chan S.W.S., Middleton D.L., Lundqvist M., Marr G.W., Higgins D.A.;
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB378701; AKS9279.1; -
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG_1like.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00409; IG; 3.
DR PROSITE; PSS0835; IG_LIKE; 2.
KW Signal.
FT SIGNAL.
SQ SEQUENCE 482 AA; 54624 MW; BDA0E8EC6172AD0B CRC64;

Query Match 8.0%; Score 271.5; DB 13; Length 482;
Best Local Similarity 25.0%; Pred. No. 2.4e-12;
Matches 111; Conservative 69; Mismatches 169; Indels 95; Gaps 19;

8 RHLLVQLALILPRAATQGNVVLGKKGDVTELTCTASOKKSIQPHNK-----NSNOIK 60
Db      12 RAVFVLQQLGL--THIMHQQQLGVGSKVILNC---KKHDKVMTKVEYDAGSSAIIIQ 66
Qy      61 ILGNQGSFLLTKGPKSLNDRADRSRLMDQGNFPLIIKNLKIENSDTYTICEVQKEEVQ 120
Db      67 ILAGK---IFKGRAPMSDRSET-----NNSKSLKVSNNRISDAGTYICEGSDRNSISL 118
Qy      121 LVFGLTANSDTHLLOQSLLTLT-LESPSSPS-----VQCRSPRGKN 162
Db      119 HAVKLTISNGYFLPGDDLELTVMKSPKSPQRPFSITLFRSHNSKRVTPVELQNETQ--- 175
Qy      163 IOGKTLVSYQLELDSDGTWTCVTLQNOKKVEFKI--DIVLAFOKAS-SIVYKKEGEV 219
Db      176 ---KVALKVKQLPRTDSGTWICMMSHSDSPSININISFNKVLVDFEKTHERVAAVDSV 232

```

Qy	220	EFSPFIATVETLUT----	SGELMMQOERASSSKSWITFDLKNKEVSKRATQDPKQM	274
Db	233	TLSTWLN-----RKIKMKEFFTQLMMQESNA-----	ITYELDFRATDGLRETK----	280
Qy	275	GKKLPLHLTLPOALP-----	OYAGSGNLTALAEAKTGKLMHOEVLNVMRATO--	321
Db	281	KRSQALLIEIPMKRDSFTEVVKIKHKLGHSGEYTCQLLYNRRYIOSKTELVMVQVSA	NP	339
Qy	322	-----LQKULTCEVWGPPTSPKML-----	SLLENKEAVSKREKVVWLNBPAGW	368
Db	340	PGPLPKAEWTLTCVSSPIPPNVHLMERVNGTQMDKRSQS--	ETKEVVKVTAVGW	397
Qy	369	QCLSDSGQVLLSEN--IKVLP	PTW 390	
Db	398	NCHLMEDNNMKLSLNTYVEAPFTW	421	
RESULT 78				
ID	0991A6	PRELIMINARY;	PRT;	484 AA.
AC	0991A6			
DT	01-JUN-2001 (TREMBLrel. 17, Created)			
DT	01-JUN-2003 (TREMBLrel. 17, Last sequence update)			
DT	01-OCT-2003 (TREMBLrel. 25, Last annotation update)			
DE	Hypotheical protein.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
CC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	Strausberg R.;			
RL	Submitted (Feb-2001) to the EMBL/Genbank/DBD databases.			
DR	EMBL; BC003495; AA03495.1; --			
DR	PIR; F33932; F33932.			
DR	HSSP; P01810; 2FBJ.			
DR	InterPro; IPR007110; IG-1like.			
DR	InterPro; IPR003006; IG_MHC.			
DR	InterPro; IPR003596; IG_v.			
DR	Pfam; PF00047; Ig_4.			
DR	SMART; SM00406; Igv.1.			
DR	PROSITE; PS50835; IG_LIKE; 4.			
DR	PROSITE; PS00290; IG_MHC; 2.			
KW	Hypotheical protein.			
SO	SEQUENCE 484 AA; 52567 MW; 8EA6A4F9BCF582FA CRC64;			
Query Match 8.0%; Score 271.5; DB 11; Length 484;				
Best Local Similarity 22.3%; Pred. No. 2.4e-12;				
Matches 142; Conservative 76; Mismatches 219; Indels 201; Gaps 26				
Qy	13	VLQALPLPAALQGNV-----	VLGKKGPTVLTGASQKKSIG--	IHKNSNQIKI 61
Db	4	VWTLFLTPAAAOISQIQIOLVQSGELKKPKGTIVKISCAASYTTTDSMHWKQAPGK	G	63
Qy	62	LGNGGSF-LTKGFSKLNDRADRSRL--	WDQGNPLIINKIIEDSDTYICEVDKKE	117
Db	64	LKMGMWVNIETGESVYADDPKGRFAPFLETSASTIHLQINMKEDTATYFCARDYD	D	123
Qy	118	VQLVFGLTANSDTHLGGOSITLTLSBPSSPSVQCRSPRKNIQGGKTLVSQLEIQ		177
Db	124	IYAM-----	DYMGQGTSTVTSSES-----	ARNP-----
Qy	178	DSGWTCTVLQNKQKVEFKIDIVLAFQKASSIVYKKGEGVSPPLATVTEKLTGSGE		237
Db	155	PA-----	LSDDVIT--	GCLHIDYF-----
Qy	238	LMMQERASSSKSWITFDLKNKEVSKRATQDPKQM	GKULPLHLTLPOALPOVAGSNL	297
Db	179	VTW-----	GKSGK-----	DITVNPFPALASGR-----
Qy	298	TL-ALEATGKLMHOEVLNVMRATQLOKNTLCEVWGPPTSPKMLSLKENKEAVSKREK		356

```

Db      210  TLPAVECPREG-----ESVKCSQVQDSNPNVQELDV-----238
Qy      357  PWWVLNPEBAGMWCGLSDSGVLLSINIKVLPWTSTFPPCPAPBAPRSCDKTHTCPELLGG 416
Db      239  -----NC-----SGFTPPRPITIPSCQ-----255
Qy      417  PSVFLPPPKPDLTLMISRTPEVTCVVVDVSHEDPEVEFNMYVVDGVENHNAKTKPREQYN 476
Db      256  PSLSLQRPALPD-LLLSQSDASITCTTLGLNRPBAV-FTW-----EPSTCKDAVQKAVQ 308
Qy      477  ST---YVWVSVLTVLHODWMLNGKEYKCKVSNKALPAPIEKTISKAKGQPREPPVYLTPPS 533
Db      309  NSCCGCVSVSLTPBCARARNMSGASFCKTVTHPE-SCGLTGTITAKVYNTPEPPVHLPPR 367
Qy      534  RDELTKNQ-VSLTCLVKGYFSPDIAVEMESNGQ---PENNYKTTTPRVLSDG--SPEFLYS 567
Db      368  SEELIATNELSLTCLVAFAPFKPEVLVWMLHGNEELSPESYLVEPFLKEPGEATTVLYVTS 427
Qy      588  KLTVDKSRNQGQANVFCSVWHEHLLHNHYTQKSLSLSTG 625
Db      428  VLRVSAETWKGQDQYSCVMGHEALPMNFTQXTIDRLSG 465

RESULT 79
088650
ID      088650      PRELIMINARY;      PRT;      120 AA.
AC      088650;
DT      01-NOV-1998 (TREMBlrel. 08, Created)
DT      01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT      01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE      T-cell surface glycoprotein CD4 (Fragment).
GN      CD4.
OS      Marmota monax (Woodchuck).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognath; Sciuridae; Sciurinae;
OC      Marmota.
CX      NCBI_TaxID=9995;
KN      [1]
RN      SEQUENCE FROM N.A.
RP      TISSUE=Periphera! blood;
RX      MEDLINE=20094937; PubMed=10627561;
RA      Guo J.T., Zhou H., Liu C., Aldrich C., Saputelli J., Whitaker T.,
RA      Batriaa M.I., Mason W.S., Seeger C.;
RT      "Apoptosis and regeneration of hepatocytes during recovery from
RT      transient hepatitisvirus infections.";
RL      J. Virol. 74:1495-1505(2000).
DR      EMBL: AF082497; AACG32621.1; -.
DR      HSSP: P01730; IWIO.
DR      GO: GO:0016020; C:membrane; IEA.
DR      GO: GO:0006955; P:immune response; IEA.
DR      InterPro: IPR000973; CD4_TcAg.
DR      InterPro: IPR007110; IG-like.
DR      PRINTS: PR00692; CD4TCANTIGEN.
PT      NON_TER      1
PT      NON_TER      1
SQ      SEQUENCE      120 AA; 13623 MW; 44967B512D520195 CRC64;

Query March      7.9%; Score 271; DB 11; Length 120;
Best Local Similarity 68.3%; Pred. No.3.6e-13;
Matches 56; Conservative 7; Mismatches 19; Indels 0; Gaps 0;

Qy      308  LHQENVLVNVRATOLQKLTCEVWGPSTSPKMLSLKLENKAKVSKREKPVWVLNPEAGM 367
Db      1  LHQGVNVLVWNVVYTHQNDLICVLAGTPPKLMLSTLKNQEAQVSKREKRIKRVNPKAGM 60
Qy      368  WQCLLSDSGVLLSINIKVLP 389
Db      61  WQCLLRDGDKVLDPQIDVDP 82

RESULT 80
091207
PRELIMINARY;      PRT;      486 AA.

```

```

AC Q91207;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA Strauberg R.;
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC010324; AAH0324.1; -
DR InterPro; IPR007110; IG_1ike.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00406; IG; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
KM Hypothetical protein.
SQ SEQUENCE 486 AA; 52682 MW; 4FEF835125DA870B CRC64;

Query Match 7.9%; Score 270.5; DB 11; Length 486;
Best Local Similarity 26.6%; Pred. No. 2,9e-12;
Matches 97; Conservative 54; Mismatches 144; Indels 69; Gaps 14;

QY 282 LTLPAALP-----QYAGSGNLTLLAEATGKLGHEVNLVVRATQLOKNTLCEV 330
DB 153 LTLPRALSDPVIIGCLIHDFPSGTMTWTW-GKSGKDTITVFPALASGGYTMSSQL 211
QY 331 WGPSPKMLSLKLENKEAVSKREKPVWVLP-PEAGMCCLSDSQVLESNIKVLP 390
DB 212 TLPA-----VECPGEGSVKCSVQHSNAPVQLDVKCS-- 243
QY 391 STPPCPAPPEPKSCDKHTHCELLGSPSVFLPPPKKDTLMISRTPEVTCVVDVSHEDP 450
DB 244 GPPPPCP-PCPPSCH-----PSLSLQRPALSD-LLSGDASITCTLNLRNPEG 290
QY 451 EVKENWYVDGVEVNAKTPREBOYNST--YRVSVLTVLHDMNGKEYCKVSKAL 507
DB 291 AV-FTW-----EPSTGKDAVQKAVQNSCGCYSVSVLPGCAERWNSGASFCTVTHPES 344
QY 508 PAPIEKTISKAKGQPREPOVYTTLPSSDELTKNQ-VSLTCLVKGFYPSDIAVEESNQ- 565
DB 345 DT-LTGITAKITVTTPFPQVHLPPPSSELAINEELVSLTCLVRAFNPKEVLRWLGNEB 403
QY 566 --PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOCGVNFSGVMEHALNHYTKSL 621
DB 404 LSPESYLVEFLPEKPEGATTTVLTSLRVSATETWQSDQYSCVGHGHALPMFTOKTID 463
QY 622 LSPG 625
DB 464 RLGG 467

```

```

RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC011703; AAH1703.1; -
DR InterPro; IPR007110; IG_1ike.
DR InterPro; IPR003597; IG_CL.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG; 4.
DR SMART; SM00407; IGCL; 3.
DR SMART; SM00406; IG; 1.
DR PROSITE; PSS0835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 2.
SQ SEQUENCE 488 AA; 53127 MW; 0E3B156E155733F0 CRC64;

Query Match 7.9%; Score 270; DB 11; Length 488;
Best Local Similarity 25.5%; Pred. No. 3,2e-12;
Matches 99; Conservative 47; Mismatches 126; Indels 116; Gaps 19;

QY 282 LTLPAALP-----QYAGSGNLTLLAEATGKLGHEVNLVVRATQLOKNTLCEV 330
DB 154 LTLPRALSDPVIIGCLIHDFPSGTMTWTW-GKSGKDTITVFPALASGGYTMSSQL 211
QY 331 WGPSPKMLSLKLENKEAVSKREKPVWVLP-PEAGMCCLSDSQVLESNIKVLP 389
DB 184 WG-----KSGKDTITVNEPPA-----LASGRYTMSSQL-TLPA 216
QY 390 WSPVPAPPEPKSCDKHTHCELLGSPSVFLPPPKKDTLMISRTPEVTCVVDVSHEDP 427
DB 217 ---VECPGEGSVKCSVQHSNAPVQLDVKCS--PPPSQPSLSLQRPAL 269
QY 428 DTLMISRTPEVTCVVDVSHEDP-VKFNWYVDGVEVNAKTPREBOYNST--YRVSV 483
DB 270 EDLLSGDASITCTLNGL--RDEGAFTW-----EPSTGKDAVQKAVQNSCGCYSVS 322
QY 484 VLVTVLHODVNLGKRYCKVSKNKLPAPIEKTISKAKGQPREPOVYTTLPSSDELTKNQ-V 542
DB 323 VLPSCAERWNSGASFCTVTHPESDT-LTGITAKITVTTPFPQVHLPPPSSELAINEELV 381
QY 543 SLTCLVKGFPYPSDIAVEESNQ--PENNYKTPPVLDSDG--SFPLYSKLTVDKSRM 597
DB 382 SLTCLVAFNPKKELVLAHMGNEBLSPEVLPKEPEGATTTVLTSLRVSATETW 441
QY 598 QGNVFSGVMEHALNHYTKSLSPG 625
DB 442 QSDQYSCVGHGHALPMFTOKTIDRLSG 469

RESULT 82
Q9DCD9 PRELIMINARY; PRT; 426 AA.
AC Q9DCD9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Adult male kidney cDNA, RIKEN full-length enriched library,
DE clone:0610041A01, full insert sequence.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=CS7BL/6J; TISSUE=Kidney;
RA MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aikawa T., Hara A., Fukunishi Y., Kono H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochiwa H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pebole G., Quackenbush J.,
RA Schirml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G.,
RA Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F.,

```

RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,  
 RA Guenrich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,  
 RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,  
 RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,  
 RA Sasaki H., Sato K., Schoenbach C., Seta T., Shibata Y., Storch K.-F.,  
 RA Suzuki H., Toyooka K., Wang K.H., Wetz C., Whitaker C., Wilming L.,  
 RA Wysshaw-Boris A., Yoshida K., Hasegawa Y., Kawaji H., Kohatsu S.,  
 RA Hayashizaki Y.,  
 RA "Functional annotation of a full-length mouse cDNA collection."  
 RT Nature 409:685-690(2001).  
 RL EMBL; AK002875; BAB22422.1; -.  
 DR EMBL; F33932; F33932.  
 DR HSSP; P01810; 2FBJ.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003597; IG\_c1.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR Pfam; PF00047; 1g; 4.  
 DR SMART; SM00407; IGc1; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 3.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 SQ SEQUENCE 426 AA, 45819 MW, 56E1275BA48F6B1 CRC64;

Query Match 7.9%; Score 269.5; DB 11; Length 426;  
 Best Local Similarity 26.6%; Pred. No. 2.9e-12;  
 Matches 97; Conservative 54; Mismatches 144; Indels 69; Gaps 14;

QY 282 LTLPOLP-----QYAGSNLTALAEKTLGHEVNLVWMARATOLQKLTCEY 330  
 DB 93 LTLPRALSDPVIIGCLIHDPSPGTNNVTW-GKSGDITTVNPPALSGGGYTWSQL 151  
 QY 331 WGPSTPKMLSLKENKAKYKREKPVWLNPEAGMOCCLSDSGQVLLSNIKVLPTW 390  
 DB 152 TLPA-----VECPGESVYKSVQHDNNAVQELDVKS--- 183  
 QY 391 STVPCCPAPPKSCDKTHTCPELLGSPVFLPPPKKDTLMISRTBEVTCVVVDVSHEDP 450  
 DB 184 GPPPCP-PCPPSCH-----PSLISQRPALSD-LLGSDASLTCTLGLNPNPG 230  
 QY 451 EVKFNMYVDCGVENNAKTKREQYNST---YRVSVLTLYLHDMDLNGEKYKKNKAL 507  
 DB 231 AV-FTW-----EPSTGDAVOKKAVONSCGYSVSVLPCCARWMSGAFKCTVTHPS 284  
 QY 508 PAPIEKTISKAKGQPREPVYTLPPSDELTKNQ-VSLTCLVKGFYPSDIAVEMESNGO- 565  
 DB 285 DT-LTGTIAKITVTNTPRQVHLPPSEBELALNELVSLTCLVAFFPKYVLRMLHGNBE 343  
 QY 566 --PENNYKTPPVLDSDG--SFPLYSKLTVDKSRWQOGVNSCSVMHEALHNHYTQKSLS 621  
 DB 344 LSPESYLVEPRLKEPGGATTVLTVSVLRVSAELMKGDQYSQVGHGALPMNFTQKTID 403  
 QY 622 LSPG 625  
 DB 404 RLSG 407

## RESULT 83

Q91XEL PRELIMINARY; PRT; 480 AA.  
 AC Q91XEL  
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)  
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein (Fragment).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Strausberg R.;  
 RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC010798; AA010798.1; -.

DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; 1g; 4.  
 DR SMART; SM00406; IG; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KW Hypothetical protein.  
 FT NON\_TER  
 SQ SEQUENCE 480 AA, 51936 MW, 20B9234BEF2B41ED CRC64;

Query Match 7.9%; Score 268.5; DB 11; Length 480;  
 Best Local Similarity 23.7%; Pred. No. 4.1e-12;  
 Matches 107; Conservative 61; Mismatches 169; Indels 115; Gaps 20;

QY 254 PDLKNEVSVKQVDPDKLQMGKPLHLTLPOLPYAGSNLTALAEKTKL---H 309  
 DB 45 FIFSNYSWVWQTEPKLEW-----VATISNSGYATHYPDSMKGRFTISR 91  
 QY 310 QEVNLVWMARATOLQKLTCEY-----EVNG-----PTSPKMLSLK 343  
 DB 92 NAQNTVLQMTSLNSBDTAVVYCTRGDYWFVWAGATTVYSSPAPRPITYLPFOA 151  
 QY 344 LENKEAKV-----SKREKPVWLN-PEAGMOCCLSDSGQVLLSNIK 385  
 DB 152 LSSDPVIIGCLIHDPSPGTNNVTWVGKSGKDTTVNFPRA-----LASGRTWSSQL- 204  
 QY 386 VLPSTVPCCPAPPKSCDKTH-----TPPELLGSPVFLPPPKK- 427  
 DB 205 TLPA-----VECPGESVYKSVQHDNNAVQELDVKS--- 257  
 QY 428 ---DTLMISRTBEVTCVVVDVSHEDP- VKENWYVDCGVENNAKTKREQYNST---Y 479  
 DB 258 RPALSDLLGSDASLTCTLNGL-RDPGAVFTW-----EPSTGDAVOKKAVONSCGY 310  
 QY 480 RVSVLTLYLHDMDLNGEKYKKNKALPAPIEKTISKAKGQPREPVYTLPPSDELTK 539  
 DB 311 SVSVLPCCARWMSGAFKCTVTHPSDT-LTGTIAKITVTNTPRQVHLPPSEBELAL 369  
 QY 540 NO-VSLTCLVKGFYPSDIAVEMESNGO--PENNYKTPPVLDSDG--SFPLYSKLTVDK 593  
 DB 370 NELSLTCLVRAFNKKEVLRVWLHGNBELSPESYLVEPRLKEPGGATTVLTVSVLR 429  
 QY 594 SRWQGVNSCSVMHEALHNHYTQKSLSLSPG 625  
 DB 430 EIKKGDDQYSCVGHGALPMNFTQKTIDRLSG 461

## RESULT 84

Q8VCV5 PRELIMINARY; PRT; 481 AA.  
 AC Q8VCV5  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC Tissue-Breast tumor;  
 RA Strausberg R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC018455; AA018455.1; -.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR InterPro; IPR003596; IG\_V.  
 DR Pfam; PF00047; 1g; 4.  
 DR SMART; SM00406; IG; 1.  
 DR PROSITE; PS50835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 2.

KW Hypothetical protein.  
 SQ SEQUENCE 481 AA: 52326 MW: 52844CS826807143 CRC64;  
 Query March 7.6% Score 259.5; DB 11; Length 481;  
 Best Local Similarity 22.9%; Pred. No. 26-11; Indels 119; Gaps 24;  
 Matches 124; Conservative 81; Mismatches 217;  
 QY 120 LTVGLTRANSPTLTLQGSGLTLTLLESP-----PGSSPVSQVCPSPGRKNI-----QGQ 166  
 Bb 6 IFLFLSGTTCVH-----SEIQLQSGPBLVPGISVAVSCASGYSIDNIYVVKSHG 61  
 QY 167 KTSLS-VSQLELQSDGTCVTCLVQKQVFEKIDIVLAFQKASSIVYKKEGEVFSFPL 225  
 Bb 62 KSLKEMIGYIDPNYNGS-----SYNQK---FK-GKATLTVDSKSTAF----- 99  
 QY 226 AFVTEKLTGSGELMWQAEBRASSKSWTFDLKNEVSVK-RVYDDPKLQMGKPLPLTLT 284  
 Bb 100 -MYLNLTLSBDSAFYICAREWYG-AMPAFWQGLTVLVASASANPTI-----YPLTL 150  
 QY 285 PQALP-----QVAGSNTLTLALAKTGKHOEVNLVVRATOLQKLTCEVWGP 333  
 Bb 151 PPLASSDPVITIGCLIHDFPBGTMVTV-GKSGDITTVFPPLASGGRITNSQLTLP 209  
 QY 334 TSPKMLSLKLENKAVSKKEKPEVYVLPAGMQLSDSGVLLSENIKVLPTWSTP 393  
 Bb 210 A-----VECPGESVSKSVQHDSPVQELDVNC-----SGP 240  
 QY 394 VPCAPAPKSCDKHTTEPELLGSPVFLFPKPDTLMISRTPEVTCVVDVSHEDPEVK 453  
 Bb 241 TPPEPITPSCQ-----PSLSLORPALED-LTSGDASITCTLNGLRNPEGAV- 287  
 QY 454 FNVVVDGVEVHNATKPREQVNST---YRVSVTLVHODMLNGKEYCKCVSKALPAP 510  
 Bb 288 FTW-----EPTGKDVAQKAVQNSCCGYSVSLPGCAERKNSGASFCTVTHPE-SGT 341  
 QY 511 IEKTSKAKQPREPOVYTLPPSRDELTKNQ-VSLTCLVKGFPSPDIAVWESNGQ---P 566  
 Bb 342 LTGTIAKVTVTFPPQVHLLPPPEBELALNELSLTCLVRAFNPKEVLRVRLHGNBELSP 401  
 QY 567 ENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOQGNVFCGVMEALHNHYTQKSLSP 624  
 Bb 402 ESYLVFEPFLKEPBGATTVTLVTSVRSVETWKGQDYSQVGHGHALPNNFTQKTIDRLS 461  
 QY 625 G-625  
 Bb 462 G-462  
 RESULT 85  
 07TMK4 PRELIMINARY; PRT; 479 AA.  
 AC 07TMK4;  
 DT 01-OCT-2003 (TREMBlrel. 25, Created)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN NCBI\_TaxID=10090;  
 RA SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Breast tumor;  
 RX MEDLINE=22386257; PubMed=12477932;  
 RA Strauberg R., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Datschenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loguclano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Boeak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,

RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Vallalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whitting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Kravinsky M.I., Skalska U., Smallus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Breast tumor;  
 RA Strauberg R.;  
 RL Submitted (AUG-2003) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; BC055905; AAH55905.1; -  
 KW Hypothetical protein.  
 SQ SEQUENCE 479 AA: 53209 MW: 532DEDD9D46D0AED CRC64;  
 Query March 7.5% Score 255; DB 11; Length 479;  
 Best Local Similarity 23.8%; Pred. No. 44-11; Indels 122; Gaps 20;  
 Matches 110; Conservative 61; Mismatches 169;  
 QY 236 GELW--WQARRASSKSWTFDLKNEV-----SVK-RVT---ODPK---LQMGK 277  
 Bb 49 GDYMMVMVROSPEKLEWVS-QIRNKRYNYETYSVSKRFTISRDSGSGVLYQNNL 107  
 QY 278 LP-----LHLLTPQALP-----QYA 292  
 Bb 108 RPEMDGIYCTVEGMDYGRGISVTVSSSARNPITYPLTLPALSSDPVITIGLIHDF 167  
 QY 293 GSGNLTALFAKTKGKHOEVNLVVRATOLQKLTCEVWPTSPKMLSLKENKAVS 352  
 Bb 168 PSGTMVTV-GKSGKDTITVFPPLASGGRITNSQLTLP----- 208  
 QY 353 KREKPVVNLNPEAGMQLSDSGVLLSENIKVLPTWSTPVPAPAPKSCDKHTCPE 412  
 Bb 209 -----VECPGESVSKSVQHDSPVQELDVNC-----SGPTPPPTTISCQ----- 250  
 QY 413 ILGSPVFLPPPKRDTLMISRTPEVTCVVDVSHEDPEVKFNVVVDGVEVHNATKPRE 472  
 Bb 251 -----PSLSLORPALED-LTSGDASITCTLNGLRNPEGAV-FTW-----EPTGKDVAQK 299  
 QY 473 EQNST---YRVSVTLVHODMLNGKEYCKSNKALPPIEKTISKAKQPREPOVYT 529  
 Bb 300 KAVQNSCGYSVSVSLPGCAERKNSGASFCTVTHPE-SGTLTGTIAKVTVNTPFPQVHL 358  
 QY 530 LPPSRDELTKNQ-VSLTCLVKGFPSPDIAVWESNGQ---PENNYKTPPVLDSDG--SF 583  
 Bb 359 LPPPEBELALNELSLTCLVRAFNPKEVLRVRLHGNBELSPESVLYFEPFLKEPBGATTV 418  
 QY 584 FLYSKLTVDKSRMOQGNVFCGVMEALHNHYTQKSLSPG 625  
 Bb 419 LVTSVLRVSAETWKGQDYSQVGHGHALPNNFTQKTIDRLSG 460  
 RESULT 86  
 08K024 PRELIMINARY; PRT; 480 AA.  
 AC 08K024;  
 DT 01-OCT-2002 (TREMBlrel. 22, Created)  
 DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Similar to expressed sequence A1893585.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 RN NCBI\_TaxID=10090;  
 RA SEQUENCE FROM N.A.  
 RC TISSUE=Breast tumor;  
 RX Strauberg R.,

RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC029188; AAH29188.1; -  
 DR InterPro: IPR003599; IG\_1  
 DR InterPro: IPR007110; IG\_1like.  
 DR InterPro: IPR003597; IG\_c1  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_v.  
 DR Pfam: PF00047; Ig\_4.  
 DR SMART: SM00409; Ig\_3.  
 DR SMART: SM00407; Ig1; 3.  
 DR SMART: SM00406; IgV; 1.  
 DR PROSITE: PS50835; IG LIKE; 4.  
 DR PROSITE: PS00290; IG\_MHC; 2.  
 SQ SEQUENCE 480 AA; 51645 MW; 8690A63C669CDBED CRC64;  
  
 Query Match 7.5%; Score 254.5; DB 11; Length 480;  
 Best Local Similarity 25.5%; Pred. No. 4.9e-11;  
 Matches 93; Conservative 55; Mismatches 147; Indels 69; Gaps 13;  
  
 QY 282 LTLPLALP-----QYAGSGNLTLLALEAKTKLHQEVNLYVMRATQLOKNTCEV 330  
 DB 147 LTLPLALSDPVIIGCLHDYFPGSTGMNVTW-GKSGKDITTVNFPALASGGRTYMSQL 205  
 QY 331 WGPSPKMLSLKLENKAKVSKREKPVWVLPNPEAGMOCILSDSGQVLLSENIKVLPTM 390  
 DB 206 TLPA-----VECPGESHVSKSVQHDSPVQELDVNC----- 236  
 QY 391 STVPCEPAPEPKSCDKTHTCPELLGSPVFLFPFKKDTLMSRTPEVTCVVVDVSHEDP 450  
 DB 237 SGTPPPPTITPSCQ-----PSLSLQRPALPD-LILGSDASITCTINGLRNPG 284  
 QY 451 EYKFNMYVDGVVHNNAKTKPREQYNST---YRVSVLTVLHQDMLNGEKYCKKYNKSL 507  
 DB 285 AV-FTW-----EPSTGKDAVQKKAQVNSCGCYSVSVLPQCAERNNGASFKCTVTHPE- 337  
 QY 508 PAPIEKTISKAKQPREPQVYTLPPSRDELTKNQ-VSLTCLVKGFPYPSDIAVEMESNQ- 565  
 DB 338 SGLTGTIAKYVNTFPQVHLLPPSEBELANELLSTLCVRAFPKPKVLRMLHGNB 397  
 QY 566 --PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOQGNVFCSVNHEALHNHYTKSL 621  
 DB 398 LSPESYLVEEPKEPGEATTVLTVSVLSAETWKQDQYSCMVGHEALPMNFTOKTID 457  
 QY 622 LSRG 625  
 DB 458 RLSG 461  
  
 RESULT 87  
 Q99M22 PRELIMINARY; PRT; 479 AA.  
 AC Q99M22  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Strauberg R.;  
 RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: BC002091; AAH02091.1; -  
 DR HSSP: P01810; 2FBJ.  
 DR InterPro: IPR007110; IG\_1like.  
 DR InterPro: IPR003006; IG\_MHC.  
 DR InterPro: IPR003596; IG\_v.  
 DR Pfam: PF00047; Ig; 4.  
 DR SMART: SM00406; IgV; 1.  
 DR PROSITE: PS50835; IG LIKE; 4.  
 DR PROSITE: PS00290; IG\_MHC; 2.

KM Hypothetical protein.  
 SQ SEQUENCE 479 AA; 51992 MW; 768B39A138918692 CRC64;  
  
 Query Match 7.4%; Score 253.5; DB 11; Length 479;  
 Best Local Similarity 25.5%; Pred. No. 5.8e-11;  
 Matches 93; Conservative 55; Mismatches 147; Indels 69; Gaps 13;  
  
 QY 282 LTLPLALP-----QYAGSGNLTLLALEAKTKLHQEVNLYVMRATQLOKNTCEV 330  
 DB 146 LTLPLALSDPVIIGCLHDYFPGSTGMNVTW-GKSGKDITTVNFPALASGGRTYMSQL 204  
 QY 331 WGPSPKMLSLKLENKAKVSKREKPVWVLPNPEAGMOCILSDSGQVLLSENIKVLPTM 390  
 DB 205 TLPA-----VECPGESHVSKSVQHDSPVQELDVNC----- 235  
 QY 391 STVPCEPAPEPKSCDKTHTCPELLGSPVFLFPFKKDTLMSRTPEVTCVVVDVSHEDP 450  
 DB 236 SGTPPPPTITPSCQ-----PSLSLQRPALPD-LILGSDASITCTINGLRNPG 283  
 QY 451 EYKFNMYVDGVVHNNAKTKPREQYNST---YRVSVLTVLHQDMLNGEKYCKKYNKSL 507  
 DB 284 AV-FTW-----EPSTGKDAVQKKAQVNSCGCYSVSVLPQCAERNNGASFKCTVTHPE- 336  
 QY 508 PAPIEKTISKAKQPREPQVYTLPPSRDELTKNQ-VSLTCLVKGFPYPSDIAVEMESNQ- 565  
 DB 337 SGLTGTIAKYVNTFPQVHLLPPSEBELANELLSTLCVRAFPKPKVLRMLHGNB 396  
 QY 566 --PENNYKTPPVLDSDG--SFPLYSKLTVDKSRMOQGNVFCSVNHEALHNHYTKSL 621  
 DB 397 LSPESYLVEEPKEPGEATTVLTVSVLSAETWKQDQYSCMVGHEALPMNFTOKTID 456  
 QY 622 LSRG 625  
 DB 457 RLSG 460  
  
 RESULT 88  
 Q9W6V7 PRELIMINARY; PRT; 487 AA.  
 AC Q9W6V7  
 DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
 DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE CD4 protein precursor.  
 GN CD4.  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 OX NCBI\_TaxID=9031;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=H.B2; TISSUE=Thymus;  
 RX MEDLINE=99218434; PubMed=10201936;  
 RA Koskinen R., Lammiakki U., Tregaskes C.A., Salomonsen J., Young J.R.,  
 Vainio O.;  
 RT "Cloning and modeling of the first nonmammalian CD4.";  
 RL J. Immunol. 162:4115-4121 (1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=B12;  
 RA Koskinen R.H., Salomonsen J., Tregaskes C.A., Young J.R.,  
 "The chicken CD4 gene has remained conserved in evolution.";  
 RT Immunogenetics 0:0-0(2002).  
 RL EMBL: Y12012; CAA72740.1; -  
 DR EMBL: AJ401223; CAC82027.1; -  
 DR InterPro: IPR007110; IG\_1like.  
 DR Pfam: PF00047; Ig; 3.  
 DR PROSITE: PS50835; IG LIKE; 2.  
 DR SIGNAL.  
 FT SIGNAL 1 28 POTENTIAL.  
 SQ SEQUENCE 487 AA; 54985 MW; B9CBA92EC9F7F45B CRC64;



Query Match 7.0%; Score 240; DB 13; Length 487;  
 Best Local Similarity 25.7%; Pred. No. 6.5e-10;  
 Matches 113; Conservative 61; Mismatches 184; Indels 82; Gaps 20;

```

10 LLLVQLALLPAATQGNKVLGKKGVLTCTA-SOKSKIQFMKN-----SNQIKLG 63
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
14 VILVQLGGLFIMAQEQI-GIAGKEVILSCAIINQDKGTWKVKYKVSSTIISFS 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
64 NQGSFLTKGPKSLNDRADRSRLMDQGNFPIIKIKLKEDSDTYICEVEDQKEVQLVF 123
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
73 KAQVFKGAPM-----THRSLSNSKSKLVSDLSLDAGIYTCACYSFVSIASHVF 125
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
124 GLTANSDTHLQGSLLTTL-ESPPGSSPS-----VQCSPPGKNIG 165
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
126 KLTISNNGHFLTNDELITLQNSHSHOPHSLKFNINDIYTTTELQESAP-KTI-- 182
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
166 GKTLVSQLELDQSGTWTCTVLQNKVFEKI-DIVLAFOKAS-SIVYKKEGEVFEFS 222
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
183 ---LKLQQLKALDGSCTWCHVSNPSINQISFDVKVIGEKERLEIYTVGNTAIIIS 239
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
223 PFLAFYBEKLTG--SGELMWAESAASSKSWITFDLKKKEVSVKRYTQDPKLQNGKPL 280
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
240 WRLNFRKI KWKEGFTGKLMEPQNTAIHELINFS-----VTHQELHRTKK-SN 288
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
281 HL-----TLPOLP--QYAGSGNTLTALEAKTGKLGHEVNLVVMRATOL----- 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
289 HIWEISEGKTGDMVDKIPVQLNHSQYRCQLEINGRTRFSRALVMQVTAIIPGPL 348
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
323 ---QKNLTCEVMGPTSPKMLSLKLEN---KEAKVSKREKPVVNLPEA-GMOC-LIS 373
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
349 SRGKMTLLCGVSPPLPSNAHLMEVRVNGTQEMKKSKQHEAKVEVNVASAPGLNCHIVE 408
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
374 DSG-QVLLSEIKYKLPWTST 392
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
409 DNNKKISLNTVEEAHVWNS 428
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 89

```

ID P70443 PRELIMINARY; PRT; 90 AA.
AC P70443;
DT 01-FEB-1997 (TREMBLrel. 02, Created)
DT 01-FEB-1997 (TREMBLrel. 02, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Mutant T-cell surface glycoprotein CD4 (Fragment).
GN CD4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BR/cdJ;
RA Rosenwasser O.A., Fairchild S.P., Tomonari K.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; U75219; AAB18133.1; -.
DR HSSP; P05540; ICID.
DR InterPro; IPR007110; IG-like.
FT NON TER
SQ SEQUENCE 90 AA; 9938 MW; 6528B707928AA685 CRC64;

```

Query Match 6.9%; Score 237; DB 11; Length 90;  
 Best Local Similarity 59.2%; Pred. No. 9.7e-11;  
 Matches 45; Conservative 14; Mismatches 17; Indels 0; Gaps 0;

```

310 QEVNLVVRATQQLQKNTICEVMGPTSPKMLSLKLENKEAKVSKREKPVVNLPEAGMWO 369
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
1 QEVNLVVRATQQLQKNTICEVMGPTSPKMLSLKLENKEAKVSKREKPVVNLPEAGMWO 60
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
370 CLUSDGQVLLSENIK 385
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 CLUSEGDKVKKMDRIQ 76
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

## RESULT 90

```

ID Q96RW7 PRELIMINARY; PRT; 5636 AA.
AC Q96RW7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hemiscentin.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RA Trent J.;
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF156100; AAK68690.1; -.
DR GO; GO:0005727; C:extrachromosomal circular DNA; IEA.
DR GO; GO:0005509; F:calcium ion binding; IEA.
DR GO; GO:0004197; F:cysteine-type endopeptidase activity; IEA.
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR InterPro; IPR00152; Asx hydroxyl_5.
DR InterPro; IPR000875; Cectropin.
DR InterPro; IPR001434; DUF11.
DR InterPro; IPR001881; EGF_Ca.
DR InterPro; IPR006209; EGF_Ca.
DR InterPro; IPR009017; GFP-like.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR00169; SHPoc_acsite.
DR InterPro; IPR000884; TSP1.
DR InterPro; IPR002035; VWF_A.
DR Pfam; PF00047; EGF_5.
DR Pfam; PF00090; tgp_1; 6.
DR SMART; SM00179; EGF_CA_7.
DR SMART; SM00408; IGc2; 43.
DR SMART; SM00209; TSP1; 6.
DR SMART; SM00327; VWA; 1.
DR TIGRFAMs; TIGR01451; B_and_repeat; 9.
DR PROSITE; PS00010; ASX_HYDROXYL; 5.
DR PROSITE; PS00268; CECTROPIN; 1.
DR PROSITE; PS01187; EGF_2; 3.
DR PROSITE; PS01187; EGF_CA; 8.
DR PROSITE; PS50835; IG_LIKE; 44.
DR PROSITE; PS00639; THIOL_PROTEASE_HIS; 1.
DR PROSITE; PS50092; TSP1; 6.
KW EGF-like domain; Immunoglobulin domain.
SQ SEQUENCE 5636 AA; 613660 MW; F000B319CED7B52C CRC64;

```

Query Match 5.8%; Score 198; DB 4; Length 5636;  
 Best Local Similarity 19.9%; Pred. No. 3.7e-05;  
 Matches 157; Conservative 105; Mismatches 252; Indels 274; Gaps 39;

```

34 GDTVELTCTASQKSKIQFMKNSNQIKLQNGSFLTKGPKSLNDRADRSRLMDQGNFP 93
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
807 GSNVTLPCVYQVPEPTIKRRLDNMPISR--PFSVSSISQARTGA----- 851
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
94 LITKLIKEDSDTYICEVEDQ---KEVQLVFGTLA-----NSDTHLQGSLLTTL 143
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
852 LFLINLWASDKGYICEAEHQFQKIQSEITVTYTGVLAPLIGISPVANVIEGOQLTLP 911
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
144 ESPPGS-----SPVQCSPPGKNIGGKTLVSQLELDQSGTWC--- 184
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
912 TLLAGNPISRRKIKASAMLQNPITYVS-----DGLITERVQLDGGSEYTCVAS 963
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
185 -----TVLQNKVFEKIDIVLAFOKAS-----SIYKKEGEVFE 221
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
964 NVAGTNKKTTSVVVHVLPTIQHQQLSTIEGIPVTLPCASGNPKPSVIMSKGELIST 1023
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

QY      222  SFPLAFYETKLTGS-GELMMQOAEKRSASSKSMV-----TFDKNKEV-----SV 263
Db      1024  S-----SAKESAGADGSSLYVVSPEGESEGEYCATATNAGYAKKRVQLTVYVRPRVEDL 1078
QY      264  KRVTDPLQMGKKLPLHLTLR-QALP-----OYAGSGNLTLA 300
Db      1079  RGLSDGRKVEIISVLAGEEVLTPCEVKSLLPRPIITWAKETQLISPSRPHTLPGSGMKIT 1138
QY      301  LEAKT-----GKLHQEVNLVYMRATOLQK-----NLTCVWGF 333
Db      1139  -ETRTSDSGMYLCVATNITAGVNTQAVKLNVHVPKIQGPKHLKVQVGQRVDICNAGST 1197
QY      334  TSPKLM-----SLKLENKAKVSKREKRVVW---LNPEAGMMQCL-----SDSQOVLLE 381
Db      1198  PLPVITWSKGSYMLVDGSHHVSNDPGLSLSDQATPSDAGIYTCVATNITAGTDEIETILH 1257
QY      382  ----SNIVL-PTWST-----FVPCBA--PEPKSCDKHTPCELLG-GPSVFL 421
Db      1258  VQEPPTVEDLEPPNYTTFQERVAHQRIEFPCKAKGTPKP-PIKMLHNGRELTLGNEPGISI 1316
QY      422  FPPPKDITLMT-SRTP-----EVTCCVVVDV-----HEDPEVX----- 453
Db      1317  L---EEGTLVLVIASVTPYDNGEYICVAANEGATTEKRYNLKVHVPVILKDEKQSVNSVL 1373
QY      454  -----FNNVVDGVEVHNKAKTKREKQVNSYSTRVVSVLVYLDHDMVNG 495
Db      1374  LNQLTNLFCVEVEGTPSPILIMMYKDVQVYTESST---IQYVNGGKILKLPFRATPED--A 1426
QY      496  KEYKCKVENKA-----LPAPIEKTISKAKQOPREHOVYTLPPSHDELTKNOVS 543
Db      1427  GRYSCKANINAGTSQKYNIDVLVPTLIIGT-----NPFKEVSV-----VLMRQVA 1472
QY      544  LTLCLVKG-FYPSEDIAVEWESNGQPPNNYKTPRPVILDSGSEFFLYSKLTVDSKRWQGNVF 602
Db      1473  LECQYKGTGTPFPD---IHHFKDQKPLFLGDVPVELLDRCQVILTKNARANDKGRYQ----- 1524
QY      603  SCSVNHQA 610
Db      1525  -CTVSNAA 1531

```

```

RESULT 91
Q90524
ID Q90524 PRELIMINARY; PRT; 268 AA.
AC Q90524;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Novel antigen receptor (Fragment).
OS Ginglymostoma cirratum (Nutree shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Galeoidea; Orectolobiformes;
OC Ginglymostomatidae; Ginglymostoma.
OX NCBI_Taxid=7801;
RN [1]
RP SEQUENCE OF 1-240 FROM N.A.
RC TISSUE=Spleen;
RX MEDLINE=95183140; Pubmed=78776689;
RA Greenberg A.S., Avila D., Hughes M., Hughes A., McKinney E.C.,
RT Flajnik M.F.;
RT "A new antigen receptor gene family that undergoes rearrangement and
RT extensive somatic diversification in sharks.";
RL Nature 374:168-173(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC TISSUE=Spleen;
RA Greenberg A.S.;
RL Submitted (DEC-1994) to the EMBL/genbank/DBJ databases.
DR EMBL; U18680; AAB48352.1; -.
DR HSSP; P01842; 7FAB.
DR GO; GO:0004872; F:receptor activity; IEA.
DR InterPro; IPR007110; IG-1-like.
DR InterPro; IPR003597; IG_C1.

```

DR	InterPro:	IPR003066;	Ig_MHC.
DR	Pfam:	PF00047; Ig_2.	
DR	SMART:	SM00407; IGcl_1.	
DR	PROSITE:	PS50835; IG_LIKE; 2.	
DR	PROSITE:	PS00290; IG_MHC; 1.	
KM	Receptor.		
FT	NON TER	268	268
SQ	SEQUENCE	268 AA; 29579 MW;	OCDE524DD6E9FDC4 CRC64;
Query Match		5.4%;	Score 186; DB 13; Length 268;
Best Local Similarity		24.3%;	Pred. No. 3.9e-06;
Matches	65; Conservative	39; Mismatches	98; Indels 66; Gaps 10

```
OY      417 PSVFLPPPKKOTLMISRPE-----WTCVVVDVSHPEDXEKMYYDGG----- 460
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      14 PYVF-----TAVDQTPQITKEGSELSINCVLADSNCAPLPYMMRKKSGSTNEE 65
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
OY      461 -----VEHNAKTPRREQYNSTYRVSVLTLYLHODWLNGKEYICKY--SNKALPAP 510
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      66 TISKGRAYEVETVNSGSKSFSLRIND-----LVV-----EDSGTYCKCYRRKKMAYDCG 113
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
OY      511 IEKT-----ISAKGQRPREPOYTL- -PSRDLETKNOVSLTCLVKGPFPEDIA 557
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      114 LEEELDMITYYGCGTGVTYVPGLIPLSPVLISHATEEQRANGVOQLVCILSIGTEPENIA 173
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
OY      558 VEMESNGOP-ENNYKITTPVLDSDFELYAKLTIVDKSRMOQGNVFSCSYMHREALHNHYT 616
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      174 VSMQKTKTITSGAFNTSPTVKTSNDFFCASHLLKVLPQEWERGSGVYSICGVSHATSSNQIR 233
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
OY      617 QRSLSLSPGLQIDETCAEAQGEIDLGM 644
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db      234 KEIRSTS-----EIAVLRLDPYVEII 255
          ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
```

RESULT 92	ID	Q8WU38	PRELIMINARY;	PRT;	573 AA.
AC	Q8WU38;				
DT	01-MAR-2002 (TrEMBLrel. 20, Created)				
DT	01-MAR-2002 (TrEMBLrel. 20, Last sequence update)				
DT	01-OCT-2003 (TrEMBLrel. 25, Last annotation update)				
DE	Hypothetical protein.				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.				
CX	NCBI_TaxID=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Tonsil;				
RA	Strausberg R.;				
RL	Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; BC021276; AAH21276.1; -.				
DR	PIR; S21205; S21205.				
DR	InterPro; IPR007110; IG-like.				
DR	InterPro; IPR003006; IG_MHC.				
DR	InterPro; IPR003596; IG_v.				
DR	Pfam; PF00047; IG_4.				
DR	SMART; SM00406; IGV; 1.				
DR	PROSITE; PS50835; IG_LIKE; 4.				
DR	PROSITE; PS00290; IG_MHC; 2.				
KW	Hypothetical protein.				
SQ	SEQUENCE 573 AA; 62967 MW; FD072344033AC530 CRC64;				

```

Query Match      5.4%  Score 185.5; DB 4; Length 573;
Best Local Similarity 21.3%; Pred. No. 1.3e-05;
Matches 153; Conservative 78; Mismatches 245; Indels 243; Gaps 36

QY      1  MNRGVPRFHLILVL---QALPLPAATQGNKVLGKGGTVELTCTAS--QKSIQPHMKX 55
Db       1  MELGISWTFLLAILKGVCEQVLVSGGGVL---QPGKSLRSCAASGTFPDYAMHWVR 57
QY      56  SNQIK-----ILGNQGSF---LTGKPSKINDPARRSLMDQGNFPLIINKLIEDS 104

```

```

Db 58 QAPGKLEWVSGISWNSGICGVADSVKGRFTIS-RDNANKSLYLQNM-----SLRAEDT 110
Qy 105 DTVICEDVDKEEYQVLLVFGLTANSDFHLQGSGLTLTLSPSSPSV-----QCRSPR 159
Db 111 ALTYCAHAGSGSYI-----GYTGMND--VMGGTTVTSAPTKADVPPIISGCHPK 162
Qy 160 GKN-----IQQGKTLVSQLELDQSGTWTCTYLQNKQKVEFKIDIVLAFQASSIVY 212
Db 163 DNSPVLACLITGYHPTSVTV-----TW-----Y 186
Qy 213 KKEGEQVEFFSP-----LAFVYEKLTGSGELMWQAE-----PASSKSNITFDLKNKE 260
Db 187 MGTSQSPQRTFPEIORRDSVYMTSSQLPQLQWRQGEYKCVQVHTASKS-----KKE 239
Qy 261 VSVKRVNODPRLQMGKPLHLTLPLQALPOVAGS---GNLLALEATKGLHQEVLVW 317
Db 240 IF--RWESPQAQAS-----SVPTQPOQAEGSLAATTAATPATRTNGRGE----- 284
Qy 318 RATOLQKRLTCEVWGPTSPKMLSLKLENKAKYKREKPVVNLPEAGMWQCLSDSGQ 377
Db 285 -----KKEKEKEGEQERET----- 299
Qy 378 VLESNIKVLFTWSTPVPCPAPEPKSCDKHTCELLIGSPVFLPPPKDQTLMIKTPPE 437
Db 300 -----KTP--ECP-----SHTQF--LG---VYLLTPAVQD-LMLRDKAT 330
Qy 438 VTGVVVDVSHEDPEVKNMWYVDG--VEVHNAKTRPEQVNSTYVSVLTVLHDDWLNGK 496
Db 331 FTCCVVSQDLKDAHL--TWEAVAGVPTGVEEGLEHNSGSGQSHRLTLPRSLMAGT 388
Qy 497 EYKCVENKALP-----APIEKTIS-KAKQPREPVQVYTLPSRDELTKNOV 542
Db 389 SVCTLNLHPSLPPEQLMALREPAQAQVYKLSLNLASSDPPEASW----- 434
Qy 543 SLTCLVKGFPYPSDIAVWESNGQPENN--YKTPPVLDSDGSFELYSKLTVDKSRWQG 599
Db 435 -LICEVSGFSPNILLMWLEDPQREVNTSGFAPARPPQSGTTFWMSVLRVPAPSPQ 493
Qy 600 NVFSCSVNHE--ALNHHYTKSLSLSPGLQDFTCAADQGLD-----GLWTT 646
Db 494 ATYTCVSHEDSRTLNL--ASRSLEVS-YLMTPLPLQSKDENDSDYTFDDVGSIMWT 549

RESULT 93
Q9D8M4 PRELIMINARY; PRT; 130 AA.
ID Q9D8M4;
AC Q9D8M4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE 1810027001RIK protein.
GN 1810027001RIK.
OS Mus musculus (Mouse).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Pancreas;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Aizawa K., Itawa Y., Nishi K., Kiyosawa H., Kondo S., Yamana K. I.,
RA Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
RA Kuehl L.M., Lewis S., Matsuo Y., Nakai D., Pezole G., Quackenbush J.,
RA Schirral L.M., Straub F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barh G.,
RA Blakes J., Botelli D., Bojunga N., Carninci P., de Bonaldo M.F.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Guelinich S., Hill D., Hofmann M., Hume D.A., Kamiya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P.,

```

```

RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L.,
RA Wyszewski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohseki S.,
RA Hayashizaki Y.,
RT "Functional annotation of a full-length mouse cDNA collection.";
RL Nature 409:685-690(2001).
DR EMBL; AK007622; BAB25142.1; -.
DR PIR; S22760; S22760.
DR HSP; P01842; TPAB.
DR MGD; MGI:1916426; 1810027001RIK.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; IG_1.
DR SMART; SM00407; IGc1.1.
DR PROSITE; PSS0835; IG_LIKE.1.
DR PROSITE; PSS00290; IG_MHC.1.
SQ SEQUENCE 130 AA; 14253 MW; 438197975E766E54 CRC64;

Query Match 5.4%; Score 183.5; DB 11; Length 130;
Best Local Similarity 36.5%; Pred. No. 2.1e-06;
Matches 38; Conservative 18; Mismatches 43; Indels 5; Gaps 3;

Qy 520 GQPR-EPQVYTLPSRDELTKNOVSLTCLVKGFPYPSDIAVWESNGQPENNYKTPPVL 577
Db 25 GQKSSPSVTLFPSSSELETNKATLVCTTIDYFGVYVDMKVDGTPVQGHETTPSK 84

Qy 578 DSDGSFELYSKLTVDKSRWQGVFSCSVNHEALNHHYTKSLS 621
Db 85 QSNKKWASVYLTTLRAWEHRSYSGQVTHE--GHVYKSLS 125

RESULT 94
Q8N355 PRELIMINARY; PRT; 234 AA.
ID Q8N355;
AC Q8N355;
DT 01-OCT-2002 (Tremblrel. 22, Created)
DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
DT 01-OCT-2003 (Tremblrel. 25, Last annotation update)
DE Hypothetical protein.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Brain;
RA Strausberg R.;
RL Submitted (APR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC028090; AAH28090.1; -.
DR PIR; S12441; S12441.
DR InterPro; IPR003599; IG.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR InterPro; IPR003596; IG_v.
DR Pfam; PF00047; IG_2.
DR SMART; SM00409; IG_2.
DR SMART; SM00407; IGc1.1.
DR SMART; SM00406; IG_1.
DR PROSITE; PSS0835; IG_LIKE.2.
DR PROSITE; PSS00290; IG_MHC.1.
KM Hypothetical protein.
SQ SEQUENCE 234 AA; 24792 MW; CC848CAEBA4A9D63 CRC64;

Query Match 5.4%; Score 183.5; DB 4; Length 234;
Best Local Similarity 25.0%; Pred. No. 4.9e-06;
Matches 55; Conservative 39; Mismatches 99; Indels 27; Gaps 7;

Qy 404 CDKTHPCPELLIGSPVFLPPPKDQTLMIKSTREVTGVVVDVSHEDPEVKFNYY----- 457
Db 14 CTGSVTSVYLTQPPSVVAP-----GQTARITCGANNISK-----SVHHYQDKPQ 60

```

```

QY 458 VDGVEVHNAKTRKPR---EEQYNSTYRVSVLTLYHODMLNGKRYCKKSNKALPAP1---- 511
DB 61 APVLVYVDSDDPSGIPERFSGNSGNTATLTISRVDAGDEAYQQLWDDSSDHPVFG 120
QY 512 EKTISAKAGQPR-EPOVYTLPPSRDELTKNOVSLTGVKGFYPSDIAYEWESNGQP-ENN 569
DB 121 GGTGKLTVLQGPKAAPSVTLFPSSSEELQANKATLVCLISDFYGGATVAMKADSPVAKG 180
QY 570 YKTPPVLDSDGSFFLYSKLTVDKSRMGOQNVFSCVMHE 609
DB 181 VETTPSKQSNMKNKAAASYSLTPQEWKSHRSYSCQVTHE 220

RESULT 95
Q99M11 PRELIMINARY; PRT; 235 AA.
ID 099M11;
AC 099M11;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN 11
RP SEQUENCE FROM N.A.
RA Strausberg R.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL: BC002129; AAH02129.1; -.
DR HSP: P01703; 7FAB
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_2.
DR SMART: SM00406; IgV_1.
DR PROSITE: PSS0835; IG_LIKE_2.
DR PROSITE: PSS00230; IG_MHC; 1.
KW Hypothetical protein.
SQ SEQUENCE 235 AA; 25403 MW; 39807BFE6782A3FB CRC64;

Query Match 5.4%; Score 183.5; DB 11; Length 235;
Best Local Similarity 24.9%; Pred. No. 5e-06;
Matches 58; Conservative 39; Mismatches 101; Indels 35; Gaps 9;

QY 407 THTCPBLGSPVFLPPKPKDITMTSRPEVTCVVDVSHEDPEVKFNMY----- 457
DB 15 TSGCAQLV-----LTQPSVST-SLGGTAKLPCKASTGNIGDSYV--NMVQGYMGRSPT 65
QY 458 --VDGVEVHNAKTRKPR---EEQYNSTYRVSVLTLYHODMLNGKRYCKKSNKALPAP 510
DB 66 NMVYGGDLRPSGVSDRSGSIDSSNSAF-----LTIQNVQADDEADYCYOSSSGIRYF 120
QY 511 EKTISAKAGQPR-EPOVYTLPPSRDELTKNOVSLTGVKGFYPSDIAYEWESNGQP-EN 568
DB 121 GGTGKLTVLQGPKAAPSVTLFPSSSEELQANKATLVCLISDFYGGATVAMKADSPVAKG 180
QY 569 NYKTPPVLDSDGSFFLYSKLTVDKSRMGOQNVFSCVMHEALHNHYTKSL 621
DB 181 GVETTPSKQSNMKNKAAASYSLTLTAKAMETHSSYSCQVTHE---GHTVEKSL 220

RESULT 96
Q96AA2 PRELIMINARY; PRT; 6620 AA.
ID 096AA2;
AC 096AA2;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Obscurin.
GN OBSCN.
OS Homo sapiens (Human).

```

```

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
RN 11
RP SEQUENCE FROM N.A.
RA Young P.W., Ehler E., Gautel M.;
RT "Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor
RT protein involved in sarcomere assembly.";
RT J. Cell Biol. 154:123-136(2001).
CC -1- SIMILARITY: CONTAINS 1 PH DOMAIN.
DR EMBL: AJ002535; CAC44768.1; -.
DR Genew; HGNC:15719; OBSCN
DR GO: GO:0005524; F:ATP binding; IEA.
DR GO: GO:0004812; F:RNA ligase activity; IEA.
DR GO: GO:0006418; P:amino acid activation; IEA.
DR InterPro: IPR003961; FN III.
DR InterPro: IPR008957; FN_III-like.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003588; Ig_c2.
DR InterPro: IPR000448; IQ_region.
DR InterPro: IPR001849; PH.
DR InterPro: IPR000219; RhogEF.
DR InterPro: IPR001452; SH3.
DR InterPro: IPR001412; trna-synt_I.
DR Pfam: PF00041; FN3; 2.
DR Pfam: PF00047; IG; 2.
DR Pfam: PF00612; IQ_1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00621; RhogEF; 1.
DR SMART: SM00060; FN3; 2.
DR SMART: SM00408; IGC2; 17.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00326; SH3; 1.
DR PROSITE: PSS0178; AA_TRNA_LIGASE_I; 1.
DR PROSITE: PSS0010; DH_2; 1.
DR PROSITE: PSS0835; IG_LIKE; 46.
DR PROSITE: PSS0096; IQ; 1.
DR PROSITE: PSS0003; PH_DOMAIN; 1.
KW Immunoglobulin domain.
SQ SEQUENCE 6620 AA; 721665 MW; C2A8EB87B284452 CRC64;

Query Match 5.3%; Score 182.5; DB 4; Length 6620;
Best Local Similarity 19.7%; Pred. No. 0.00072;
Matches 135; Conservative 93; Mismatches 269; Indels 187; Gaps 28;

QY 29 VLGKGDYVELTCTASOKKSIQFMKNSNQIKLNGQSFLLTKPSKLNDRADSRSLMD 88
DB 1818 VCGELGVTTLTLCELS-PACAEVVRGCGTQPRVGRKRFQMAVEGPR----- 1863
QY 89 QGNFLLIKNLKIEDSDTYICEVEDQKEVQLLVFGLTANSDTHLLQGQSLTTLSSPG 148
DB 1864 ----SLTVLGLAEDAGEVYCESRDDHTSAQLTV----- 1893
QY 149 SSPVQCSPPRAKNIQGGKTLVSQLELQDSTWTCTYLQNKRYEFKIDIVLAF---- 204
DB 1894 ---SV-----PRVYKMSG--LSTVAEEGEGATFCVSPS-----DVAVWFRDGA 1936
QY 205 --QKASIVYKKEGQVERSPFLAFTVEKLTGSGELMQAERASSKSWITFDLKNKEVS 262
DB 1937 LLOPSEKPAISQSGA----SHSLTISDLVLEAGQITVEAEGASSAA-----LKRREAP 1987
QY 263 V--KRVTDPKLQMGKGLPLHLTLPLQALPQVAGSGNLTALBAKTKGLHOE---VNLVVM 317
DB 1988 VLFKKLEPQVVEERSSVTLVELTRPWPBELRTNATATLARGKVEIHAEGARHLVLH 2047
QY 318 RATIOLOKN-LTCEWGPSPSKMLSLKLENKAKYSKREKPPVWVNLPEAGMOCCLSDSG 376
DB 2048 NVGFADRGFCET---PDDKTQAKLTVMRQVRLVR-----GLQAVEAREQ 2092
QY 377 OVLE---SNIKVLPWSTPVPVCPAPBPKSCDKHTCPBLGSP--SVFLPPKPKDT-L 430

```

```

Db      2093 TATREVOLSHADVGSWTR-----DGLRFQGGPTCHLAVRGPMHTLTLGLRPEDSGL 2145
Qy      431 MISR-----TPEVTCVVDVSHDEDEYKFNWVVDGEV 463
Db      2146 MVEFAEGVHTSARLVLTLPVFSRPLQDVVTTEKEKVTLECELSRNVVDVRLKDDVEL 2205
Qy      464 HNAKTREREQYNSTYRVSVLTVDHMDLNGKEYCKVSNKALPAPL----- 511
Db      2206 RAGKTMALAAQ-----GACRSLLTYRCEPADQGVYVDADHDAQSSASVKKQGRYYTLTYR 2260
Qy      512 -----EKTISKAKGQRPQVYTLPPSSDELT--KNQVSLTCLYKGFYPSD 555
Db      2261 RVLAEDAGEIQFVLENASERQAQVKEPLVTLVRLPDKIAMEGHGVLEQVQV---RAS 2317
Qy      556 IAVWESNG---QENNYKTPPLVLDSDGFPLYSKLTVDSKRNQOGNVSQSV--MHEA 610
Db      2318 AQVWFKGSELOQGPVYE-----LVSDG--LYRKLLISDVHAEDEDTYTCADGVKTS 2369
Qy      611 LHHNYTOKSLSLSPGLQDETCAE 634
Db      2370 AQFVEEQSITIVRGLQ-DVTWME 2392

RESULT 97
O8ISF3 PRELIMINARY; PRT; 2693 AA.
AC O8ISF3;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 301kDa_2 protein.
GN ISOF.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307;
RA Flaherty D., Garnett K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions."
RL J. Mol. Biol. 323:533-549(2002).
DR 1 EMBL; AY130758; AAN61521.1;
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR GO; GO:0004688; F:protein amino acid phosphorylation; IEA.
DR InterPro; IPR003961; FN_III.
DR InterPro; IPR003598; IG_II.
DR InterPro; IPR007110; IG_II-like.
DR InterPro; IPR003598; IG_c2.
DR InterPro; IPR000719; Prot_Kinase.
DR InterPro; IPR002290; Ser_Thr_Kinase.
DR InterPro; IPR008271; Ser_Thr_Kin_AS.
DR InterPro; IPR001245; Tyr_Kinase.
DR Pfam; PF00041; fn3; 2.
DR Pfam; PF00047; Ig; 19.
DR Pfam; PF00069; pkinase; 1.
DR ProDom; PD000001; Proc_kinase; 1.
DR SMART; SM00060; FN3; 1.
DR SMART; SM00408; IGc2; 19.
DR SMART; SM00220; S_TKc; 1.
DR SMART; SM00219; TYRc; 1.
DR PROSITE; PS00835; IG_LIKE; 18.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PSS0011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PSS0108; PROTEIN_KINASE_ST; 1.
SQ SEQUENCE 2693 AA; 299094 MW; 1C84AD2011E8391A CRC64;

```

Query Match 5.3%; Score 179.5; DB 5; Length 2693;

```

Best Local Similarity 19.8%; Pred. No. 0.00034;
Matches 133; Conservative 106; Mismatches 258; Indels 175; Gaps 32;

Qy      28 VLGKGGDTVELTCTASQKKSIOFHWKNSQIQLNGQSFLLTKGSKLNDRAISRSLM 87
Db      1100 VVLTAGEITATFTQSANPAQVW-----LHNGKALQOTKSNYKTRLF 1144
Qy      88 DQGNFPLIKNLIEISDPTICEVEDQKEEV-----OLLVGLTRANS----- 129
Db      1145 DDNTATLIVENYVDELQGITAVANNQFGDVHTSAQLTISGSAKTIASLPFFIIEIKP 1204
Qy      130 DTHLQOSLTLLTPSPGSSP-----SVQCRSP-----GKNIQGGKTLVSQ 173
Db      1205 KINWVG--ATLSIQADLNGSPLEPVWLKDNSELVESDRILQKCGVWQV-----LVVRD 1258
Qy      174 LEIQDSGTWTCTYLQNKQKFEKIDIVLAFOQASSTLVKKEGEQVYF-----S 222
Db      1259 VGLDEGTITITANEKGIQNTQVSV--TSKVEYKEKKEKVEKKDEGKKKGRPG 1315
Qy      223 FPL--AFTVEKLT-----GSGELMWQAEASSSKSWITF-DLKNEVSVK----- 264
Db      1316 LPRPSGASKTEQVTMAFPDAPSBGPAUSVEYERCPQREWVSGSTSLLEIKGLTPNT 1375
Qy      265 ----RVTPDKLQMGKPLPLHTLPOA---LPQVAGSNLTALBAKTGKLHQEVNLV 316
Db      1376 EYIFRVAGKKNQKQGLGEMSEMTSLTKTASVQQAQF-----TISPQK-----IIA 1420
Qy      317 MRATLOKNTLCVGEWPTSPKMLSLKENKAKVSKREK-----PVWVLN---PEA 365
Db      1421 NRDEPE--IAVFGSTPFP-----SVKMYKENIQIVDEKIDVATTSTSSILMLKQSEN 1474
Qy      366 GMMQCLL-SDSGVLESNIKVLPWTSTPPCAPAPKSDCKHTCEBELLGGSVLPFP 424
Db      1475 GTFNCLLENLGGASASCOVTT--FNKPSLQSTPDHSLERN-----LVPT 1518
Qy      425 KPR----DTLMSRTPEVTCVVDVSHDEDEYKFNWVVDGEVHNAKTREREQYNSTYR 480
Db      1519 LQKALNNESQAQOQIMLTIRISRS-----ESTVAMFKDERIESAGRYLSDDKSNHK 1574
Qy      481 VSVLTVLHODMNGKEYCKVSNKALPAPIETKISKANGQPREPOVYTLPPS----- 533
Db      1575 LV-CHAVQSQD--TGK-YRCVVTNKYGAESQCNVA-----EDVTKFIAPFSATLSD 1624
Qy      534 RDLTKNQSVLTLVYGFYSDIAVWESNGQENNYKTPPLVLDSDGFPLYSKLTVDK 593
Db      1625 STALLGHNTLLECKVGSAPAE--VSWTDGERISTTRIRIQTDQNGN---CKLSISK 1678
Qy      594 SRWQGNVFSQS 605
Db      1679 AESDDMGVYCS 1690

RESULT 98
O8ISF4 PRELIMINARY; PRT; 2708 AA.
AC O8ISF4;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE 301kDa_1 protein.
GN ISOF.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditioidea;
OC Rhabditidae; Pelodertinae; Caenorhabditis.
OX NCBI_Taxid=6239;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22269627; PubMed=12381307;
RA Flaherty D., Garnett K., Shmeleva N., Tang X., Mercer K.,
RA Borodovsky M., Benian G.;
RT "Titins in Caenorhabditis elegans with Unusual Features: Coiled-coil
RT Domains, Novel Regulation of Kinase Activity and Two New Possible
RT Elastic Regions."

```

RL J. Mol. Biol. 323:533-549 (2002).  
 DR EMBL: AY130758; RAN61520.1; -.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO: GO:0004713; F:protein-tyrosine kinase activity; IEA.  
 DR GO: GO:0004668; F:protein-amino acid phosphorylation; IEA.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR007110; IG\_II.  
 DR InterPro: IPR003598; IG\_C2.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR008271; Ser\_thr\_pkin\_AS.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00041; fn3; 2.  
 DR Pfam: PF00047; Ig; 19.  
 DR Pfam: PF00069; pkinase; 1.  
 DR ProDom: PD000001; Prot\_kinase; 1.  
 DR SMART: SM00060; FN3; 1.  
 DR SMART: SM00408; IGC2; 19.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR SMART: SM00219; TyKc; 1.  
 DR PROSITE: PS00835; IG\_LIKE; 18.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 DR SEQUENCE 2708 AA; 300729 MW; 28BEEB856FA571 CRC64;

Query Match 5.3%; Score 179.5; DB 5; Length 2708;  
 Best Local Similarity 19.8%; Pred. No. 0.00034;  
 Matches 133; Conservative 106; Mismatches 258; Indels 175; Gaps 32;

QY 28 VILGKGDVVELTCTSSQKSIQPHKNSQIKILNGSFLTKGSKLNDRAEDRSRLM 87  
 DB 1115 VILKTAGEATFTFQSYVAPAAQVW-----LHNGALQOTKSNYTRLF 1159  
 QY 88 DQGNFLIINKLIEISDFYICEVEDQKEV-----OLVGLTPANS----- 129  
 DB 1160 DQNTATLVIENVDDELGGITTAANNQFGDVHTSAQITSSGSAKILASLPFIETELK 1219  
 QY 130 DTHLLQGQSITLTLESPPGSSP-----SVQCRSP-----GKNIQGGKTLVSQ 173  
 DB 1220 KINVEG--ATLSIQADLNGSPIPEVWLKDNSELVEDRIQMKCGVNVQ-----LLVND 1273  
 QY 174 LEIQDSGTWTCTYLQMKKEFEKIDIVLAFOQASIVYKKEBOJEF-----S 222  
 DB 1274 VGLDEGTITTTNENKGIKQNTESV--TTSKEVKEKKEKVEKDEGKKRGPRG 1330  
 QY 223 FPL--AFVTEKLT-----GSGELMQAERASSKSWITP-DIKNKEVSVK----- 264  
 DB 1331 LPRPSGASKEQVTMAFDAPSEGPADSYEVERCPCPQRENVSCGSKSLELEIKGLTPNT 1390  
 QY 265 ----RVTOPDKLQMGKKLPLHLTLPOA---LPQVAGSGLTLALAKTGKLGHEVNLV 316  
 DB 1391 EYIFRAVAGKNGQGLGEMSEWSTLTKIASVQAPQF-----TISPSK----- 1435  
 QY 317 MRATLOKNTLTCVWGPSTPKMLSLKLENKAKVSKREK-----PWVNL-----PEA 365  
 DB 1436 NRDEFE--IAVEFSGTPP-----SVKMYKENQIIVDEKIDVATTSTSIINLKSGEEN 1489  
 QY 366 GMYQCLL-SPSGCVLLESNIKVLPTWSTPVCAPAPKSCDKHTCPBELLGSGSVFLPP 424  
 DB 1490 GTNCLLENLQASASQVTT--FNKPSLSTSTDSHSLER-----LVPL 1533  
 QY 425 KPR----DTLMISRTPEVTCVVVDVSHEDPEVKFNNYVVDGVEVNAKTKRREQQSTYR 480  
 DB 1534 LQKALNNESAQAQOQIMLTCRISRS--ESTVAWFKDERIESAGRYELSDSKSNHK 1589  
 QY 481 VSVVLVYLVHODVWNGKRYKRVGNKALPAPIETIKSAKQPREPOVYTLPPS----- 533  
 DB 1590 LV-CHAVQSD--TGK-YRCVWVNNKYGYASECNVAN-----EDYTKFLAPSPSATLSD 1639  
 QY 534 RDELTKNQVSLTCTLVGFGFYPSDIAVEWESNGQDENNYKTPPYLDSGSEFLYSKLTVDK 593

DB 1640 STAILGHNTILECKVEGSPAPE--VSWTKDGERISTRRIRQTOPDENGN-----CKLSISK 1693  
 QY 594 SRMQGQNVFSGS 605  
 DB 1694 AESDDMGVYVCS 1705

## RESULT 99

Q8MNSO PRELIMINARY; PRT; 2780 AA.  
 AC Q8MNSO;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical protein.  
 GN F12F3.2.  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
 OC Rhabditidae; Pelodertinae; Caenorhabditis.  
 OC NCBI\_Taxid=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA "The sequence of C. elegans cosmid F12F3."  
 RT Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Fulcon B., Wohlmann P.;  
 RT "The sequence of C. elegans cosmid F12F3."  
 RT Submitted (JAN-1997) to the EMBL/Genbank/DBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Waterston R.;  
 RT Submitted (MAY-2002) to the EMBL/Genbank/DBJ databases.  
 RL EMBL: U80022; AAM29672.1; -.  
 DR WormRep; F12F3.2a; CB30753.  
 DR GO: GO:0005524; F:ATP binding; IEA.  
 DR GO: GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO: GO:0004713; F:protein-tyrosine kinase activity; IEA.  
 DR GO: GO:0016740; F:transferase activity; IEA.  
 DR GO: GO:0004668; F:protein-amino acid phosphorylation; IEA.  
 DR InterPro: IPR003962; FNIII\_subd.  
 DR InterPro: IPR003961; FN\_III.  
 DR InterPro: IPR008957; FN\_III-like.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003598; IG\_C2.  
 DR InterPro: IPR000719; Prot\_kinase.  
 DR InterPro: IPR002290; Ser\_thr\_kinase.  
 DR InterPro: IPR008271; Ser\_thr\_pkin\_AS.  
 DR InterPro: IPR001245; Tyr\_pkinase.  
 DR Pfam: PF00041; fn3; 2.  
 DR Pfam: PF00047; Ig; 19.  
 DR Pfam: PF00069; pkinase; 1.  
 DR PRINTS: PR00014; FNTYPERII.  
 DR ProDom: PD000001; Prot\_kinase; 1.  
 DR SMART: SM00060; FN3; 2.  
 DR SMART: SM00408; IGC2; 19.  
 DR SMART: SM00220; S\_TKc; 1.  
 DR SMART: SM00219; TyKc; 1.  
 DR PROSITE: PS00835; IG\_LIKE; 18.  
 DR PROSITE: PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE: PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE: PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW Hypothetical protein; Immunoglobulin domain; Repeat;  
 KW Transferrase.  
 QW SEQUENCE 2780 AA; 308646 MW; 046D057107B935C1 CRC64;

Query Match 5.3%; Score 179.5; DB 5; Length 2780;

[illegible]

SN	[2]	SEQUENCE FROM N.A.
RP		STRAIN=Bristol NZ;
RC		Fulton B., Wohlmann P.;
RT		"The sequence of C. elegans cosmid F12f3."
RL		Submitted (JAN-1997) to the EMBL/GenBank/DDBJ databases.
RN	[3]	
RP		SEQUENCE FROM N.A.
RC		STRAIN=Bristol NZ;
RA		Waterston R.;
RL		Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR		EMBL: U80022; AAM29673.1; -
DR		MormRep; F12F3.2b; CE30754.
DR		GO; GO:0005524; P:ATP binding; IEA.
DR		GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.
DR		GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.
DR		GO; GO:0006740; F:transferase activity; IEA.
DR		GO; GO:0006468; P:protein amino acid phosphorylation; IEA.
DR		InterPro; IPR003962; FNIII_subd.
DR		InterPro; IPR003961; FN_III.
DR		InterPro; IPR007110; Ig-like.
DR		InterPro; IPR003598; Ig_c2.
DR		InterPro; IPR000719; Prot_kinase.
DR		InterPro; IPR008271; Ser_Chk_pkin_AS.
DR		InterPro; IPR01245; Tyr_kinase.
DR		Pfam; PF00047; fn3; 2.
DR		Pfam; PF00069; pkinase; 1.
DR		PRINTS; PR00014; FMYPERII.
DR		ProDom; PD000001; Proc_kinase; 1.
DR		SMART; SMO0060; FN3; 2.
DR		SMART; SMO0408; IGc2; 19.
DR		SMART; SMO0220; S_TKC; 1.
DR		SMART; SMO0219; TYR_KC; 1.
DR		PROSITE; PS50835; IG_LIKE; 18.
DR		PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR		PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR		PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW		Hypothetical protein; ATP-binding; Immunoglobulin domain; Repeat; Transferase
SQ		SEQUENCE 2808 AA; 311979 MW; 02AFED0AE06FE12 CR664;
Query Match	5.3%; Score 179.5; DB 5; Length 2808;	
Best Local Similarity	19.8%; Pred. No. 0.00036;	
Matches 133; Conservative 106; Mismatches 258; Indels 175; Gaps 32;		
QY	28	VTLGKGDVYEELTTTAAQSQKSIGFHMKNSNOIKILNQSGSLVTGSPSKANDRADRSRLW 87
Db	1187	VLTMTAETATPTFTQSYANPAQAQVM-----LNHGKALQQTKSNKYTRLTF 12311
QY	88	DQGPFLLIRKLKIEDSDTYICEVEDQKEEY---OLLVFGLTANS----- 129
Db	1232	DNDTAITLVIEVDDELGGITYTAVANNQFGDVHTSAQLTISGEAKKIATAASLPYTIIELEK 12911
QY	130	DTHLLOQSLLTLESPPGSSP-----SVQCRSPR-----GKNIOGKTLSVSQ 173
Db	1292	KINNVEG--ATLSIQADLNGSPIPEVVWLMDNSSELVESDRIQMKCDGVNQ---LVRD 1345
QY	174	LELDOSGTWTCTYLONQKVEFKIDIVLAFQKASSIVYKKEGQVEF-----S 222
Db	1346	VGLEDESTLYITIAENEGKITROMREVAV---TSKSEVKKEKKKVEKDGKKKPGRP 1402
QY	223	FPL---AFVTEKLT-----SGGLMWQAERASSSKSMITF-DLKNEKESVYK----- 264
Db	1403	LIPRPSGASKTEQIVMAFDAPSEGAGDSIYEVERRCPDQREVWSCGSTSLBELRIKGLTPNT 1462
QY	265	----RVYQDPFLQMGKCLPLHLEPQA-----LPQYASGNULTLALEAKTGKHQEVALLV 316
Db	1463	EYIRVAGKKKQGGIGEWSEMTSTILTAKTSVGQAQPF-----TISPQK-----IIA 1507
QY	317	MRATOLQKNLTCEVWGFTSPKMLSLKENKAQVSRREK-----PWVLN---PEA 365

```

Db 1508 NRDEFE--IAVEFSGTPTP---SVKWKYKENVQIVPEKIDVATSTSSILNLKSQEN 1561
QY 366 GMMQCLL-SDSGVLLIESNIKVLPWTSTVPCCPAPBPSCDKHTHCPELLGGPSVFLPP 424
Db 1562 GFENCLIEENELGQASASCQVTI---FNKPASLOSTPDHSLERN-----LVPT 1605
QY 425 KPK---DTLMTSRTEPVTCVVVDVSHEDPEVKFNNYVVDGVEVHNAKTKPREEQYNSTYR 480
Db 1606 LQKALNESAQAGQOIMLTGRISRS---ESTVAMFKDDERIESAGRYELSSDKKSNHK 1661
QY 481 VVSVLTVLHODMLNGREYKCKVSNKALPAPIEKTISKAKGQPREPQVYTLPPS----- 533
Db 1662 LV-CHAVQSD--TGR-YRCVVTNKKYGAISECNVAV-----EDVTKFIAPSFATLSD 1711
QY 534 RDELTKQVSLTCLVKGFIPSDIAVEMESNGQPENNYKTTTPVLDSGSPFLYSKLTVDK 593
Db 1712 STAILGHNITLLECKVEGSPAPE--VSMTKDGERISTTRIRIOTODENGN---CKLSISK 1765
QY 594 SRWQGNVFCSS 605
Db 1766 AESDDMGVYCS 1777

```

Search completed: August 3, 2004, 13:13:49  
 Job time : 53.4955 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 12:50:44 ; Search time 8.8991 Seconds  
(without alignments)  
3791.557 Million cell updates/sec

Title: SEQ07  
Perfect score: 3414  
Sequence: 1 MNRGVPPRHLLVLQLALP.....DETCAEAQDELGLWTTDP 648

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 125 summaries

Database : SwissProt\_42:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2015	59.0	458	1	CD4_HUMAN
2	1991	58.3	458	1	CD4_PANTR
3	1844	54.0	458	1	CD4_MACFU
4	1843	54.0	458	1	CD4_MACMU
5	1840	53.9	458	1	CD4_MACPA
6	1835	53.7	458	1	CD4_MACNE
7	1818	53.3	458	1	CD4_CERAE
8	1717	50.3	397	1	CD4_BRYPA
9	1707	50.0	397	1	CD4_CERTO
10	1572.5	46.1	457	1	CD4_SALSC
11	1211.5	35.5	330	1	GC1_HUMAN
12	1150.5	33.7	459	1	CD4_RABIT
13	1143	33.5	290	1	GC3_HUMAN
14	1135	33.2	463	1	CD4_CANFA
15	1115.5	32.7	326	1	GC2_HUMAN
16	1105.5	32.4	327	1	GC4_HUMAN
17	1000.5	29.3	457	1	CD4_RAT
18	993.5	29.1	457	1	CD4_MOUSE
19	938.5	27.5	398	1	GC3M_MOUSE
20	928.5	27.2	393	1	GC1M_MOUSE
21	910	26.7	323	1	GC_RABIT
22	902.5	26.4	399	1	GCAM_MOUSE
23	879.5	25.8	329	1	GC2_MOUSE
24	879.5	25.8	405	1	GCBM_MOUSE
25	840.5	24.6	329	1	GC3_MOUSE
26	829	24.3	326	1	GC1_RAT
27	826.5	24.2	333	1	GCB_RAT
28	822.5	24.1	329	1	GCB_RAT
29	814.5	23.9	324	1	GC1_MOUSE
30	801.5	23.5	330	1	GCAM_MOUSE
31	792	23.2	332	1	GC_A_RAT
32	788	23.1	335	1	GCAB_MOUSE
33	783.5	22.9	336	1	GCB_MOUSE

34	368.5	10.8	421	1	EPC_MOUSE
35	363.5	10.6	429	1	EPC_RAT
36	359.5	10.5	476	1	MUCM_MOUSE
37	358.5	10.5	454	1	MUC_HUMAN
38	356	10.4	428	1	EPC_HUMAN
39	355.5	10.4	479	1	MUCM_RABIT
40	352.5	10.3	391	1	MUCB_HUMAN
41	348	10.2	455	1	MUC_MOUSE
42	346	10.1	457	1	MUC_GUNNU
43	346	10.1	458	1	MUC_RABIT
44	342.5	10.0	454	1	MUC_MESAU
45	340	10.0	450	1	MUC_CANFA
46	306	9.0	299	1	ALC_RABIT
47	303	8.9	461	1	HVCM_HETER
48	301.5	8.8	438	1	HVCS_HETER
49	300	8.8	446	1	MUC_CHICK
50	297.5	8.7	438	1	HVC2_HETER
51	293.5	8.6	340	1	ALC2_HUMAN
52	289.5	8.5	393	1	HVC3_HETER
53	282.5	8.3	353	1	ALC1_GORGO
54	282.5	8.3	353	1	ALC1_HUMAN
55	270	7.9	370	1	HVCI_HETER
56	261.5	7.7	481	1	MUCM_ICTPU
57	251.5	7.4	344	1	ALC_MOUSE
58	199.5	5.8	513	1	SHS1_MOUSE
59	184	5.4	4391	1	PCBM_HUMAN
60	177.5	5.2	105	1	LAC1_MOUSE
61	177.5	5.2	3707	1	PCBM_MOUSE
62	174.5	5.1	6332	1	UN89_CAEEL
63	174	5.1	739	1	VCA1_RAT
64	172	5.0	104	1	LAC2_RAT
65	171	5.0	506	1	SHS1_BOVIN
66	165	4.8	702	1	CEAS_HUMAN
67	163	4.8	104	1	LAC3_MOUSE
68	162.5	4.8	213	1	ILL1_HUMAN
69	162.5	4.8	1493	1	NEOL_MOUSE
70	162	4.7	1906	1	KML5_CHICK
71	161	4.7	104	1	LAC2_MOUSE
72	161	4.7	258	1	HB2D_PIG
73	160.5	4.7	739	1	VCA1_HUMAN
74	159	4.7	105	1	LAC_HUMAN
75	156	4.6	105	1	LAC_PIG
76	153.5	4.5	105	1	LAC5_MUSSP
77	153	4.5	105	1	LAC_RABIT
78	153	4.5	106	1	KACB_RABIT
79	152.5	4.5	847	1	CD272_HUMAN
80	151.5	4.4	383	1	DTG_HUMAN
81	150	4.4	106	1	KACB_RAT
82	149.5	4.4	1447	1	DCC_MOUSE
83	148.5	4.3	1257	1	CAML_HUMAN
84	148	4.3	106	1	KAC_HUMAN
85	147	4.3	739	1	VCA1_MOUSE
86	146	4.3	104	1	LAC1_RAT
87	146	4.3	387	1	SRB2_HUMAN
88	145.5	4.3	103	1	LAC_CHICK
89	145	4.2	106	1	KACB_RAT
90	145	4.2	261	1	HB2C_PIG
91	145	4.2	509	1	SHS1_RAT
92	145	4.2	837	1	NCM2_MOUSE
93	144.5	4.2	555	1	C166_CARAU
94	143.5	4.2	503	1	SHS1_HUMAN
95	143.5	4.2	1260	1	SHS1_MOUSE
96	143	4.2	106	1	KAC_MOUSE
97	142.5	4.2	398	1	SRB1_HUMAN
98	141.5	4.1	1447	1	DCC_HUMAN
99	141	4.1	6885	1	SNE2_HUMAN
100	140	4.1	268	1	HB2X_HUMAN
101	140	4.1	273	1	ZDOB_HUMAN
102	140	4.1	1259	1	CAML_RAT
103	139.5	4.1	2012	1	DSOB_HUMAN
104	139	4.1	273	1	CD4_PANTR
105	138.5	4.1	564	1	C166_BRARE
106	138.5	4.1	1197	1	CAML_BRARE

P06336	mus musculus
P01855	rattus norv
P01873	mus musculus
P01871	homo sapien
P01854	homo sapien
P04221	oryctolagus
P04220	homo sapien
P01872	mus musculus
P20768	suncus muri
P03988	oryctolagus
P06337	mesocricetu
P01874	canis famli
P23088	oryctolagus
P23087	heterodontu
P01875	gallus gall
P23085	heterodontu
P01877	homo sapien
P23086	heterodontu
P20758	gorilla gor
P01876	homo sapien
P23084	heterodontu
P23735	ictalurus p
P01878	mus musculus
P97797	m. proteini-t
P98160	homo sapien
P01843	mus musculus
Q05793	mus musculus
P29534	rattus norv
O01761	caenohabdi
P20767	rattus norv
O46631	bos taurus
P06731	homo sapien
P01845	mus musculus
P15814	homo sapien
P97798	mus musculus
P11799	gallus gall
P01844	mus musculus
P15980	sus scrofa
P19320	homo sapien
P01842	homo sapien
P01846	sus scrofa
P20765	mus spreus
P01847	oryctolagus
P01835	oryctolagus
P20273	homo sapien
P01880	homo sapien
P70211	mus musculus
P32004	homo sapien
P01834	homo sapien
P29533	mus musculus
P20766	rattus norv
O91976	rattus norv
P20763	gallus gall
P01833	rattus norv
P15983	sus scrofa
P97710	r. proteini-t
O35136	mus musculus
P01837	mus musculus
O00241	homo sapien
P33146	homo sapien
O8wx10	homo sapien
P05558	homo sapien
P13765	homo sapien
Q05695	rattus norv
O60469	homo sapien
P18467	pan troglod
Q90460	brachydanio
Q90478	brachydanio

ID	CD4_HUMAN	STANDARD	PRT	458 AA
107	138	4.0	997	1 SCPI RAT
108	137.5	4.0	761	1 NCAG_HUMAN
109	136.5	4.0	1348	1 VGR2_COTUA
110	136	4.0	261	1 HB24_HUMAN
111	135.5	4.0	105	1 LACS_MOUSE
112	134.5	3.9	261	1 HB22_HUMAN
113	134	3.9	1369	1 NFAS_CHICK
114	133	3.9	837	1 NCW2_HUMAN
115	132.5	3.9	261	1 HB21_HUMAN
116	132.5	3.9	1240	1 NFAS_MOUSE
117	132	3.9	231	1 HB21_CHICK
118	131.5	3.9	848	1 NCAG_HUMAN
119	131	3.8	261	1 HB23_HUMAN
120	130	3.8	1277	1 CAML_FICGRU
121	130	3.8	1302	1 NRG_DROME
122	130	3.8	1694	1 IGAG_HAEIN
123	129	3.8	1541	1 IGAL_HAEIN
124	128.5	3.8	1240	1 NFAS_RAT
125	128	3.7	1141	1 MYPS_HUMAN

## ALIGNMENTS

## RESULT 1

ID	CD4_HUMAN	STANDARD	PRT	458 AA
AC	P01730			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	01-NOV-1988 (Rel. 09, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Les-3).			
OS	CD4.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_TaxID=9606;			
RN	[1]			
RX	SEQUENCE FROM N.A.			
RX	MEDLINE=85254948; PubMed=2990730;			
RA	Maddon P.J., Littman D.R., Godfrey M., Maddon D.E., Chess L., Axel R.;			
RA	"The isolation and nucleotide sequence of a cDNA encoding the T cell surface protein T4: a new member of the immunoglobulin gene family."			
RL	Cell 42:93-104(1985).			
RN	[2]			
RN	REVISION TO 26.			
RX	MEDLINE=89028665; PubMed=3263213;			
RA	Littman D.R., Maddon P.J., Axel R.;			
RA	"Corrected CD4 sequence."			
RL	Cell 55:541-541(1988).			
RN	[3]			
RN	SEQUENCE FROM N.A.			
RX	MEDLINE=96303695; PubMed=8723724;			
RA	Aseati-Lari M.A., Muzny D.M., Lu J., Lu F., Lillley C.E., Spanos S., Malley T., Gibbs R.A.;			
RA	"A gene-rich cluster between the CD4 and triosephosphate isomerase RT genes at human chromosome 12p13."			
RL	Genome Res. 6:314-326(1996).			
RN	[4]			
RN	SEQUENCE FROM N.A. AND VARIANT TRP-265.			
RX	MEDLINE=91216786; PubMed=1708753;			
RA	Hodge T.W., Saaso D.R., McDougal J.S.;			
RA	"Humans with OKT4-epitope deficiency have a single nucleotide base RT change in the CD4 gene, resulting in substitution of TRP240 for ARG240."			
RL	Hum. Immunol. 30:99-104(1991).			
RN	[5]			
RN	SEQUENCE FROM N.A.			
RC	TISUE=Pancreas;			
RX	MEDLINE=22388257; PubMed=12477932;			
RA	Straubeberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,			

RA	Klausener R.D., Collins F.S., Wagner L., Shermen C.M., Schuler G.D., Alteschul R.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K., Hopkins R.P., Jordan H., Moore T., Max S.I., Wang J., Heish F., Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L., Stapleton M., Soares M.B., Donald M.F., Casavanti T.L., Scheetz T.E., Brownstein M.J., Uddin T.B., Toshiyuki S., Carninci P., Prange C., Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Miliaty S.J., Bosak S.A., McGowan P.J., McKernan K.J., Malek J.A., Gunaratne P.H., Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W., Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A., Fahy J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A., Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G., Blakesley R.W., Touchman J.W., Green B.D., Dickson M.C., Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Rodriguez Y.S.N., Krzywicki M.I., Skalska U., Smalls D.E., Schermer A., Schein J.E., Jones S.J.M., Maitra M.A.;
RA	"Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences."
RT	Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN	[6]
RN	SEQUENCE OF 28-424 FROM N.A.
RC	TISUE=Blood;
RX	MEDLINE=93049640; PubMed=1425921;
RA	Fomsgaard A., Hirsch V.M., Johnson P.R.;
RA	"Cloning and sequences of primate CD4 molecules: diversity of the RT cellular receptor for simian immunodeficiency virus/human immunodeficiency virus."
RL	Eur. J. Immunol. 22:2973-2981(1992).
RN	[7]
RN	SEQUENCE OF 26-394.
RX	MEDLINE=90078232; PubMed=2592374;
RA	Carr S.A., Hemling M.E., Folea-Maesserman G., Sweet R.W., Anumula K., Barr J.R., Huddleston M.J., Taylor P.;
RA	"Protein and carbohydrate structural analysis of a recombinant soluble CD4 receptor by mass spectrometry."
RT	J. Biol. Chem. 264:21286-21295(1989).
RN	[8]
RN	X-RAY CRYSTALLOGRAPHY (2.4 ANGSTROMS) OF 26-208.
RX	MEDLINE=91061881; PubMed=1701030;
RA	Wang J., Yan Y., Garrett T.P., Liu J., Rodgers D.W., Garlick R.L., Tarr G.E., Husein Y., Reinherz E.L., Harrison S.C.;
RA	"Atomic structure of a fragment of human CD4 containing two immunoglobulin-like domains."
RL	Nature 348:411-418(1990).
RN	[9]
RN	X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF 26-208.
RX	MEDLINE=91061882; PubMed=2247146;
RA	Ryu S.-E., Kwong P.D., Truneh A., Porter T.G., Arthos J., Rosenberg M., Dai X., Xuong N.-H., Axel R., Sweet R.W., Hendrickson W.A.;
RA	"Crystal structure of an HIV-binding recombinant fragment of human CD4."
RL	Nature 348:419-426(1990).
RN	[10]
RN	X-RAY CRYSTALLOGRAPHY (3.9 ANGSTROMS) OF 26-388.
RX	MEDLINE=97311402; PubMed=9168119;
RA	Wu H., Kwong P.D., Hendrickson W.A.;
RA	"Dimeric association and segmental variability in the structure of human CD4."
RL	Nature 387:527-530(1997).
RN	[11]
RN	PALMITOYLATION.
RX	MEDLINE=92317088; PubMed=1618861;
RA	Crise B., Rose J.K.;
RA	"Identification of palmitoylation sites on CD4, the human immunodeficiency virus receptor."
RT	J. Biol. Chem. 267:13593-13597(1992).
CC	-I- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.
CC	-I- SUBUNIT: Associates with p56-lck.
CC	-I- SUBCELLULAR LOCATION: Type I membrane protein.
CC	-I- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC	-I- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.

```

CC -1- DATABASE: NAME=PROW, NOTE=CD guide CD4 entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prow/cd/cd4.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M12807; AAA5572.1; -
DR EMBL; U47924; AAB51309.1; -
DR EMBL; M35160; AAA16069.1; -
DR EMBL; BC025782; AAB25782.1; -
DR PIR; A90872; RHMUT4.
DR PDB; 1CDH; 30-APR-94.
DR PDB; 1CDI; 30-APR-94.
DR PDB; 3CD4; 31-OCT-93.
DR PDB; 1CDJ; 01-APR-97.
DR PDB; 1CDU; 01-APR-97.
DR PDB; 1CDV; 01-APR-97.
DR PDB; 1CDB; 12-MAR-97.
DR PDB; 1WIO; 07-JUL-97.
DR PDB; 1WIP; 07-JUL-97.
DR PDB; 1WIO; 07-JUL-97.
DR PDB; 1G9M; 27-DEC-00.
DR PDB; 1G9N; 27-DEC-00.
DR PDB; 1GCI; 19-AUG-98.
DR PDB; 1JL4; 19-SEP-01.
DR GlycoSiteDB; P01730; -
DR GlycoSiteDB; P01730; -
DR HGN; 186940; -
DR GO; GO:0042101; C:T-cell receptor complex; NAS.
DR GO; GO:0015026; F:coreceptor activity; NAS.
DR GO; GO:0015029; F:internalization receptor activity; NAS.
DR GO; GO:0042289; F:MHC class II protein binding; NAS.
DR GO; GO:0004888; F:transmembrane receptor activity; NAS.
DR GO; GO:0006955; F:immune response; NAS.
DR GO; GO:0009405; F:pathogenesis; NAS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; NAS.
DR GO; GO:0030217; P:T-cell differentiation; NAS.
DR GO; GO:0045058; P:T-cell selection; NAS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; NAS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; IG_Like.
DR Pfam; PFO0047; IG_2.
DR PRINTS; PRO0692; CD4TCANTIGEN.
DR SMART; SMO0406; IGv_1.
DR PROSITE; PS50835; IG_Like; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure;
KW Polymorphism.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT VARIANT 265 265
R -> W (in OKT4-negative populations).

```

```

FT STRAND 27 32 /FTID=VAR_003906.
FT TURN 33 34
FT STRAND 37 39
FT TURN 44 45
FT STRAND 51 55
FT TURN 56 57

Query Match 59.0%; Score 2015; DB 1; Length 458;
Best Local Similarity 99.5%; Pred. No. 3,4e-121;
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

1 MNRGVPFPHLLVQLALPAATGQNKVYGGKGDVETLCTASQKSIQFHWGNSQIK 60
1 MNRGVPFPHLLVQLALPAATGQNKVYGGKGDVETLCTASQKSIQFHWGNSQIK 60
1 ILGNQGSFLTKGPKNDNRADSRSLWDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120
61 ILGNQGSFLTKGPKNDNRADSRSLWDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120
121 LVFGTLNSTHLLQGSQSLTLTLSPGSSPSVQCRRPRGKNTGGKTLSSVQLELDSG 180
121 LVFGTLNSTHLLQGSQSLTLTLSPGSSPSVQCRRPRGKNTGGKTLSSVQLELDSG 180
121 TWCTVQNKQKVEFKIDIVLAFOKASIVYKKEGQVEFSPPLATVEKLTGSGELMW 240
181 TWCTVQNKQKVEFKIDIVLAFOKASIVYKKEGQVEFSPPLATVEKLTGSGELMW 240
181 TWCTVQNKQKVEFKIDIVLAFOKASIVYKKEGQVEFSPPLATVEKLTGSGELMW 240
241 QABRASSKSMITFDLKNKEVSRYVTQDPKLGKGLPLHLTLPOALPOYAGSGNLTLA 300
241 QABRASSKSMITFDLKNKEVSRYVTQDPKLGKGLPLHLTLPOALPOYAGSGNLTLA 300
301 LEAKTGLHOBVNLVWRATQLOKNTCEWGPSPPLMLSLKLENKAVSKREKVVW 360
301 LEAKTGLHOBVNLVWRATQLOKNTCEWGPSPPLMLSLKLENKAVSKREKVVW 360
361 LNPBAGMWQCLSDSGQVLLSNKVLPTWSTVP 394
361 LNPBAGMWQCLSDSGQVLLSNKVLPTWSTVP 394

RESULT 2
CD4_PANTR STANDARD; PRT; 458 AA.
AC P16004;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Pan troglodytes (Chimpanzee).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Pan.
OX NCBI_TaxID=9598;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=9018264; Pubmed=2107024;
RX Camerini D., Seed B.;
RT "A CD4 domain important for HIV-mediated syncytium formation lies
RT outside the virus binding site.";
RL Cell 60:747-754(1990).
RN [2]
RP SEQUENCE OF 26-424 FROM N.A.
RA MEDLINE=93049640; Pubmed=1425921;
RX Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
receptor interaction. May regulate T-cell activation.

```

```

CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31135; AAA35407.1; -.
DR EMBL; X73323; CAAS1749.1; -.
DR PIR; B32722; RWC2T4.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TCRg.
DR InterPro; IPR007110; Ig_V.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT CONFLICT 42 42
FT CONFLICT 191 191
FT SEQUENCE 458 AA; 51057 MW; A7C3AC8A5257D3AD CRC64;
Query Match 58.3%; Score 1991; DB 1; Length 458;
Best Local Similarity 98.5%; Pred. No. 1.2e-119;
Matches 388; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

```

```

Db 181 TWTCVLAQNGKVFEPKIDIVLAFQKASIVYKKEGQVFSFPLAFTVEKLTGSGELMW 240
Qy 241 QAEPRASSSSSWITFDLAKKEVSVKVTDPPKIQMGKPLPLHLTLPALQVYAGSSNLTLA 300
Db 241 QAEPRASSSSSWITFDLAKKEVSVKVTDPPKIQMGKPLPLHLTLPALQVYAGSSNLTLA 300
Qy 301 LEAKTKLHQEAVNLVVMRATQLOKNU/TCCEWGPSPKMLSLKLEKAKVSKREKPVWY 360
Db 301 LEAKTKLHQEAVNLVVMRATQLOKNU/TCCEWGPSPKMLSLKLEKAKVSKREKPVWY 360
Qy 361 LNPEAGMQLSDSGQVLLLESINIKVLPWTSPV 394
Db 361 LNPEAGMQLSDSGQVLLLESINIKVLPWTSPV 394
RESULT 3
ID CD4_MACFU STANDARD; PRT; 458 AA.
AC P79184;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Les-3).
GN CD4.
OS Macaca fuscata fuscata (Japanese macaque).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9543;
RN [1]
RP SEQUENCE FROM N.A.
RA Hashimoto O., Tatsumi M.;
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D63348; BAA09672.1; -.
DR HSSP; P01730; 1MBR.
DR GO; GO:0042101; C:T-cell receptor complex; ISS.
DR GO; GO:0015026; F:coreceptor activity; ISS.
DR GO; GO:0042289; F:MHC class II protein binding; ISS.
DR GO; GO:0006955; P:immune response; ISS.
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P:T-cell differentiation; ISS.
DR GO; GO:0045058; P:T-cell selection; ISS.
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TCRg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig; 2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
CYTOPLASMIC (POTENTIAL).

```

```

FT DOMAIN 26 125 IG-LIKE V-TYPE.
FT DOMAIN 126 203 IG-LIKE C2-TYPE 1.
FT DOMAIN 126 203 IG-LIKE C2-TYPE 2.
FT DOMAIN 318 374 IG-LIKE C2-TYPE 3.
FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 41 109 BY SIMILARITY.
FT DISULFID 155 184 BY SIMILARITY.
FT DISULFID 328 370 BY SIMILARITY.
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).
SQ SEQUENCE 458 AA; 50828 MM; 76B3E7EF08185535 CRC64;

Query Match 54.0%; Score 1844; DB 1; Length 458;
Best Local Similarity 90.4%; Pred. No. 2.7e-110;
Matches 356; Conservative 17; Mismatches 21; Indels 0; Gaps 0;

QY 1 MNRGVPRHLLLVQLALLPAATOGNKKVLLGKGDVETLTCTASQKKSIOFHMKNNOIK 60
D 1 MNRGIPRHHLLLVQLALLPAATOGKVKVLLGKGDVETLTCTASQKKNTOFHMKNNOIK 60
QY 61 ILGNQGSFLTKGPKSLNDRADSRSLMDQGNPFLIKNKLIEDSTYICEVEDQKEEYOL 120
D 61 ILGIGSFLTKGPKSLNDRADSRSLMDQGSFMIKIKLKIEDSTYICEVENKKEEVEL 120
QY 121 LVFGLTNSDTHLLOQGLTTLTSSPPSSPVOCRSRGNIOGKTLVSQHELDQSG 180
D 121 LVFGLTNSDTHLLEQGLTTLTSSPPSSPVCRSGKGNIOGRTISVPLERDQSG 180
QY 181 TWCTCTVONOKKVEFKIDIVLAFQKASSIVYKKEGEVRSFPPLAFVETLTSSEGLMW 240
D 181 TWCTCTSDQKTEFEKIDIVLAFQKASSIVYKKEGEVRSFPPLAFVETLTSSEGLMW 240
QY 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKPLQNGKKPLHLTLPOALPOYAGSGNTLA 300
D 241 QAERASSSKSWITFDLKNKEVSVKRVTPDKPLQNGKKPLHLTLPOALPOYAGSGNTLA 300
QY 301 LEATGKLHGVNLVWMAATOLQKLTCEWNGPSPKMLSLKLENKAKSKREKPVW 360
D 301 LEATGKLHGVNLVWMAAFOELTEWNGPSPKMLSLKLENKATVSKAKAVW 360
QY 361 LNPEAGMOCLLSDSGQVLLSNNIKVLPFTWSTPV 394
D 361 LNPEAGMOCLLSDSGQVLLSNNIKVLPFTWSTPV 394

RESULT 4
CD4_MACMU STANDARD; PRT; 458 AA.
ID CD4_MACMU
AC P16003; Q29617;
DT 01-APR-1990 (Rel. 14, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
T4/Leu-3).
CN CD4.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=90182664; PubMed=2107024;
RA Cameroni D., Seed B.;
RT "A CD4 domain important for HIV-mediated syncytium formation lies
RT outside the virus binding site.";
RL Cell 60:747-754(1990).
RN (2)
RP SEQUENCE FROM N.A.
RC TISSUE=Thymocytes;
RA Hashimoto O., Tateuchi M.;

```

```

RT "Molecular cloning and expression of macaque CD4s.";
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 28-424 FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus.";
RL Eur. J. Immunol. 22:2973-2981(1992).
RN [4]
RP SEQUENCE OF 107-192 FROM N.A.
RX MEDLINE=98320644; PubMed=9656488;
RA Harris E.E., Disotell T.R.;
RT "Nuclear gene trees and the phylogenetic relationships of the
RT mangabeys (Primates: Papionini).";
RL Mol. Biol. Evol. 15:892-900(1998).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (by similarity).
CC -1- SUBCELLULAR LOCATION: Type 1 membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M31134; AAA36838.1; -
DR EMBL; D63347; BAA09671.1; -
DR EMBL; X73326; CAA51752.1; -
DR EMBL; AF057385; AAC52129.1; -
DR HSRP; P01730; 1MBR.
DR GO; GO:0042101; C: T-cell receptor complex; ISS.
DR GO; GO:0015026; F: coreceptor activity; ISS.
DR GO; GO:0042289; F: MHC class II protein binding; ISS.
DR GO; GO:0006955; P: immune response; ISS.
DR GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P: T-cell differentiation; ISS.
DR GO; GO:0045058; P: T-cell selection; ISS.
DR GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TcAg.
DR InterPro; IPR007110; Ig_Like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IG; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
FT CONFLICT 42 42
FT CONFLICT 62 62
L -> S (IN REF. 3).

```

```

FT CONFLICT 67 67 L -> S (IN REF. 2).
FT CONFLICT 169 169 I -> L (IN REF. 2).
FT CONFLICT 191 191 K -> N (IN REF. 3).
FT CONFLICT 248 248 S -> P (IN REF. 2).
FT CONFLICT 265 265 R -> Q (IN REF. 3).
FT CONFLICT 349 349 A -> T (IN REF. 2).
SQ SEQUENCE 458 AA; 50884 MW; 8BB80339FAFEC808 CRC64;

Query Match 54.0%; Score 1843; DB 1; Length 458;
Best Local Similarity 90.4%; Pred. No. 3,1e-110;
Matches 356; Conservative 17; Mismatches 21; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLVQLALPAATQGNKVVYGGKGDVETLCTASOKKSIOFHKNKSNQIK 60
DB 1 MNRGIFPRHLVLVQLALPAVTOGKKVYGGKGDVETLCTASOKKNTQFHKNKSNQIK 60
QY 61 ILGNOSFLLTKGSPKLNDRADSRSLMDQGNFPLIKNLKIENSDTYICEVEDQKEEVL 120
DB 61 ILGIQGFLLTKGSPKLNDRADSRSLMDQGFMIINLKIENSDTYICEVENKKEEVL 120
QY 121 LVFGLTANSDTHLLEQOSLTLLTLESPGSSPSVQCRSPGKNIOGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLEQOSLTLLTLESPGSSPSVQCRSPGKNIOGKRTISVPLERDQSG 180
QY 181 TWCTVLOQKQKVEFKIDIVLAFOKASIVYKKEGOVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWCTVSQDKQKVEFKIDIVLAFOKASIVYKKEGOVEFSPPLAFTLEKLTGSGELMW 240
QY 241 QAERASSKSWITFDLNKKEVSVKRYTOPDKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300
DB 241 QAERASSKSWITFDLNKKEVSVKRYTOPDKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300
QY 301 LEAKTGKLEQEVNLVYMRATQLOKNTLCEWGPSTPKMLSLKLENKAVSKREKPVVW 360
DB 301 LEAKTGKLEQEVNLVYMRATQLOKNTLCEWGPSTPKMLSLKLENKAVSKREKPVVW 360
QY 361 LNPEAGMWOCCLSDSGOVLLSNIKVLPWTSTPV 394
DB 361 LNPEAGMWOCCLSDSGOVLLSNIKVLPWTSTPV 394

RESULT 5
CD4_MACPA STANDARD; PRT: 458 AA.
ID CD4_MACPA
AC P79185;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OC NCBI_Taxid=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Thymocytes;
RA Tatsuni M., Yabe M., Yamada Y.K.;
RL Submitted (FBI-1997) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

```

```

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
EMBL: D63349; BAA09673.1; -.
DR HSSP; P01730; 1MR.
DR GO; GO:0042101; C: T-cell receptor complex; ISS.
DR GO; GO:0015026; F: coreceptor activity; ISS.
DR GO; GO:0042289; F: MHC class II protein binding; ISS.
DR GO; GO:0005955; P: immune response; ISS.
DR GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P: T-cell differentiation; ISS.
DR GO; GO:0045058; P: T-cell selection; ISS.
DR GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4 TCAG.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 458
FT DOMAIN 26 396
FT TRANSMEM 397 418
FT DOMAIN 419 458
FT DOMAIN 26 125
FT DOMAIN 126 203
FT DOMAIN 204 317
FT DOMAIN 318 374
FT CARBOHYD 42 42
FT CARBOHYD 296 296
FT CARBOHYD 325 325
FT DISULFID 41 109
FT DISULFID 155 184
FT DISULFID 328 370
FT LIPID 419 419
FT LIPID 422 422
SQ SEQUENCE 458 AA; 50872 MW; 9105479F85C6FF7 CRC64;

Query Match 53.9%; Score 1840; DB 1; Length 458;
Best Local Similarity 90.4%; Pred. No. 4,8e-110;
Matches 356; Conservative 16; Mismatches 22; Indels 0; Gaps 0;

QY 1 MNRGVPFRHLVLVQLALPAATQGNKVVYGGKGDVETLCTASOKKSIOFHKNKSNQIK 60
DB 1 MNRGIFPRHLVLVQLALPAVTOGKKVYGGKGDVETLCTASOKKNTQFHKNKSNQIK 60
QY 61 ILGNOSFLLTKGSPKLNDRADSRSLMDQGNFPLIKNLKIENSDTYICEVEDQKEEVL 120
DB 61 ILGIQGFLLTKGSPKLNDRADSRSLMDQGFMIINLKIENSDTYICEVENKKEEVL 120
QY 121 LVFGLTANSDTHLLEQOSLTLLTLESPGSSPSVQCRSPGKNIOGKTLVSQLELDQSG 180
DB 121 LVFGLTANSDTHLLEQOSLTLLTLESPGSSPSVQCRSPGKNIOGKRTISVPLERDQSG 180
QY 181 TWCTVLOQKQKVEFKIDIVLAFOKASIVYKKEGOVEFSPPLAFTVEKLTGSGELMW 240
DB 181 TWCTVSQDKQKVEFKIDIVLAFOKASIVYKKEGOVEFSPPLAFTLEKLTGSGELMW 240
QY 241 QAERASSKSWITFDLNKKEVSVKRYTOPDKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300
DB 241 QAERASSKSWITFDLNKKEVSVKRYTOPDKLQMGKKLPLHLTLPALPOYAGSGNLTLLA 300
QY 301 LEAKTGKLEQEVNLVYMRATQLOKNTLCEWGPSTPKMLSLKLENKAVSKREKPVVW 360
DB 301 LEAKTGKLEQEVNLVYMRATQLOKNTLCEWGPSTPKMLSLKLENKAVSKREKPVVW 360
QY 361 LNPEAGMWOCCLSDSGOVLLSNIKVLPWTSTPV 394
DB 361 LNPEAGMWOCCLSDSGOVLLSNIKVLPWTSTPV 394

```

RESULT 6  
CD4\_MACNE STANDARD; PRT; 458 AA.  
ID CD4\_MACNE  
AC 008340; P79196;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 15-JUL-1998 (Rel. 36, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Leu-3).  
GN CD4.  
OS Macaca nemestrina (pig-tailed macaque).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Macaca.  
OC Cercopithecoidea; Macaca.  
OX NCBI\_TaxID=9545;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hashimoto O., Tatsumi M.;  
RL Submitted (FEB-1997) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE OF 28-424 FROM N.A.  
RC TISSUE=Blood;  
RX MEDLINE=93049640; PubMed=1425921;  
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;  
RT "Cloning and sequences of primate CD4 molecules: diversity of the cellular receptor for simian immunodeficiency virus/human immunodeficiency virus."  
RL Eur. J. Immunol. 22:2973-2981(1992).  
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell receptor interaction. May regulate T-cell activation.  
CC -1- SUBUNIT: Associates with p56-lck (by similarity).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.ebi.ac.uk/announcements> or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL; D63346; BAA09670.1; -;  
DR EMBL; X73325; CAAS1751.1; -;  
DR HSSP; P01730; 1WBR.  
DR GO; GO:0042101; C:T-cell receptor complex; ISS.  
DR GO; GO:0015026; F:coreceptor activity; ISS.  
DR GO; GO:0042289; F:MHC class II protein binding; ISS.  
DR GO; GO:0006955; P:immune response; ISS.  
DR GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.  
DR GO; GO:0030217; P:T-cell differentiation; ISS.  
DR GO; GO:0045058; P:T-cell selection; ISS.  
DR GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.  
DR InterPro; IPR000973; CD4\_TcAg.  
DR InterPro; IPR007110; Ig-Like.  
DR InterPro; IPR003596; Ig\_V.  
DR Pfam; PF00047; Ig\_2.  
DR PRINTS; PR00692; CD4TCANTIGEN.  
DR SMART; SMO0406; IGV\_1.  
DR PROSITE; PS50835; IG\_LIKE\_1.  
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell; Immune response; Repeat; Signal; Lipoprotein; Palmitate.  
FT SIGNAL 1 25  
FT CHAIN 26 458  
FT DOMAIN 26 396  
FT TRANSMEM 397 418  
FT DOMAIN 419 458  
FT DOMAIN 26 125  
FT DOMAIN 126 203  
FT DOMAIN 204 317  
FT DOMAIN 318 374  
IG-LIKE C2-TYPE 3.  
IG-LIKE C2-TYPE 2.  
IG-LIKE C2-TYPE 1.  
IG-LIKE C2-TYPE 2.  
IG-LIKE C2-TYPE 3.

FT CARBOHYD 42 42 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 296 296 N-LINKED (GLCNAC. . .) (BY SIMILARITY).  
FT CARBOHYD 325 325 N-LINKED (GLCNAC. . .) (BY SIMILARITY).  
FT DISULFID 41 109 BY SIMILARITY.  
FT DISULFID 155 184 BY SIMILARITY.  
FT DISULFID 328 370 BY SIMILARITY.  
FT LIPID 419 419 S-palmitoyl cysteine (By similarity).  
FT LIPID 422 422 S-palmitoyl cysteine (By similarity).  
FT CONFLICT 57 57 D -> N (IN REF. 2).  
FT CONFLICT 91 91 C -> H (IN REF. 2).  
FT CONFLICT 105 105 N -> D (IN REF. 2).  
FT CONFLICT 113 113 N -> E (IN REF. 2).  
FT CONFLICT 302 302 D -> E (IN REF. 2).  
FT CONFLICT 349 349 T -> A (IN REF. 2).  
SQ SEQUENCE 458 AA; 50905 MW; 751A9BA2C8B3EE16 CRC64;  
  
Query Match 53.7%; Score 1835; DB 1; Length 458;  
Best Local Similarity 89.8%; Pred. No. 1e-109;  
Matches 354; Conservative 19; Mismatches 21; Indels 0; Gaps 0;  
  
QY 1 MNRGVPFRHLLVQLALLPAATQGNRVYIGKGDVVELTCTASOKKSIQPHKNSQIK 60  
DB 1 MNRGIPFRHLLVQLALLPAATVQGGKGVIGKGDVVELTCNASKKNTQPHKNSQIK 60  
  
QY 61 ILGQSGFLTKGSPKLNDRASRSLMDQGNFPIIKNLKIEPDSFYICEVDQKEEVL 120  
DB 61 ILGQSGFLTKGSPKLNDRASRSLMDQCFSMITIKNLKIEPDSFYICEVDQKEEVL 120  
  
QY 121 LVFGLTANSDTHLLQGGSLTLLTSPGSSPSVQCRSPRKNIOGKTLTSSVSOLEQDSG 180  
DB 121 LVFGLTANSDTHLLQGGSLTLLTSPGSSPSVQCRSPRKNIOGKTLTSSVPOLEQDSG 180  
  
QY 181 TWICTVQONQKVEFKDIYVLAFOKASSIVYKKEGQVEFSPPLATVEKLTGSGGLW 240  
DB 181 TWICTVQONQKVEFKDIYVLAFOKASSIVYKKEGQVEFSPPLATVEKLTGSGGLW 240  
  
QY 241 QAEBASSKSMITFDLKNKESVYKRVYQDPKLMGKKPLPHLTLPOLPOYAGSGNLTIA 300  
DB 241 QAEBASSKSMITFDLKNKESVYKRVYQDPKLMGKKPLPHLTLPOLPOYAGSGNLTIA 300  
  
QY 301 LEAKTGLHGEVNLVVRATQLOKNTLCEVWGPSTPLMLSLKLENKAVSKREKPVW 360  
DB 301 LDKTGLHGEVNLVVRATQLOKNTLCEVWGPSTPLMLSLKLENKAVSKREKPVW 360  
  
QY 361 LNPBAGMOCILSDSGQVLLSNIKVLPTSTPV 394  
DB 361 LNPBAGMOCILSDSGQVLLSNIKVLPTSTPV 394  
  
RESULT 7  
CD4\_CERAE STANDARD; PRT; 458 AA.  
ID CD4\_CERAE  
AC 008338; 002805; 077593; 028217;  
DT 01-FEB-1995 (Rel. 31, Created)  
DT 15-MAR-2004 (Rel. 43, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Leu-3).  
GN CD4.  
OS Cercopithecus aethiops (Green monkey) (Grivet).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea; Cercopithecoidea.  
OC Cercopithecoidea; Cercopithecoidea.  
OX NCBI\_TaxID=9534;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Hashimoto O., Tatsumi M.;  
RL Submitted (JUL-1996) to the EMBL/GenBank/DBJ databases.  
RC TISSUE=Blood;  
RX MEDLINE=93049640; PubMed=1425921.

RA Fomsgaard A., Hirsch V.M., Johnson P.R.:  
 RT "Cloning and sequences of primate CD4 molecules: diversity of the  
 RT cellular receptor for simian immunodeficiency virus/human  
 RT immunodeficiency virus.";  
 RL Eur. J. Immunol. 22:2973-2981(1992).  
 RN [3]  
 RP SEQUENCE OF 28-424 FROM N.A.  
 RX TISSUE=Peripheral blood;  
 RA MEDLINE=98017879; Pubmed=9379478;  
 RA Fomsgaard A., Mueller-Trutwin M.C., Diop O., Hansen J., Mathiot C.,  
 RA Corbet S., Barte-Sinoussi F., Allan J.S.;  
 RT "Relation between phylogeny of African green monkey CD4 genes and  
 RT their respective simian immunodeficiency virus genes.";  
 RL J. Med. Primatol. 26:120-128(1997).  
 RN [4]  
 RP SEQUENCE OF 107-192 FROM N.A.  
 RX MEDLINE=98320644; Pubmed=9656488;  
 RA Harris E.E., Disotell T.R.;  
 RT "Nuclear gene trees and the phylogenetic relationships of the  
 RT mangabeys (primates: Papionini).";  
 RL Mol. Biol. Evol. 15:892-900(1998).  
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell  
 CC receptor interaction. May regulate T-cell activation.  
 CC -1- SUBUNIT: Associates with p56-lck (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL: D86589; BAA1312.1; -;  
 DR EMBL: X73322; CAA51748.1; -;  
 DR EMBL: AF001226; AAB60873.1; -;  
 DR EMBL: AF001228; AAB60875.1; -;  
 DR EMBL: AF057380; AAC25124.1; -;  
 DR HSSP: P01730; IWIQ.  
 DR GO: GO:0042101; C:T-cell receptor complex; ISS.  
 DR GO: GO:0015026; F:coreceptor activity; ISS.  
 DR GO: GO:0042289; F:MHC class II protein binding; ISS.  
 DR GO: GO:0006955; P:immune response; ISS.  
 DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.  
 DR GO: GO:0030217; P:T-cell differentiation; ISS.  
 DR GO: GO:0045058; P:T-cell selection; ISS.  
 DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.  
 DR InterPro: IPR000973; CD4\_TCSA.  
 DR InterPro: IPR007110; Ig\_Like.  
 DR InterPro: IPR003596; Ig\_v.  
 DR Pfam: PF00047; Ig; 2.  
 DR PRINTS: PRO0692; CD4TCNTGEN.  
 DR SMART: SM00406; IGV; 1.  
 DR PROSITE: PS50835; IG\_Like; 1.  
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;  
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.  
 FT SIGNAL 1 25  
 FT CHAIN 26 458  
 FT DOMAIN 26 396  
 FT TRANSMEM 397 418  
 FT DOMAIN 419 458  
 FT DOMAIN 26 125  
 FT DOMAIN 126 203  
 FT DOMAIN 204 317  
 FT DOMAIN 318 374  
 FT CARBOHYD 42 42  
 FT CARBOHYD 281 281  
 FT CARBOHYD 296 296  
 FT CARBOHYD 325 325  
 FT DISULFID 41 109  
 BY SIMILARITY.

FT DISULFID 155 184 BY SIMILARITY.  
 FT DISULFID 328 370 BY SIMILARITY.  
 FT LIPID 419 439 S-palmitoyl cysteine (By similarity).  
 FT LIPID 422 422 S-palmitoyl cysteine (By similarity).  
 FT CONFLICT 46 46 K -> N (IN REF. 2 AND 3; AAB60875).  
 FT CONFLICT 59 59 I -> T (IN REF. 3 AND 3; AAB60873).  
 FT CONFLICT 115 115 K -> E (IN REF. 1).  
 FT CONFLICT 165 165 G -> V (IN REF. 3; AAB60873 AND 4).  
 FT CONFLICT 200 200 M -> V (IN REF. 2 AND 3).  
 FT CONFLICT 227 227 F -> L (IN REF. 3; AAB60873).  
 FT CONFLICT 271 271 K -> E (IN REF. 3; AAB60873).  
 FT CONFLICT 281 281 N -> H (IN REF. 3; AAB60873).  
 SQ SEQUENCE 458 AA; 51158 MW; FCS23D2EDD1F72E7 CRC64;  
 Query Match 53.3%; Score 1818; DB 1; Length 458;  
 Best Local Similarity 89.3%; Pred. No. 1.2e-108; Indels 0; Gaps 0;  
 Matches 352; Conservative 17; Mismatches 25;  
 QY 1 MNRGVPRHLVLVQLALPAATQGNKVLGKKGDTVELCTPASKSIQPHKNSNQIK 60  
 DB 1 MNMGIFRHLVLVQLALPAATQGNKVLGKKGDTVELCTNASKOTTQPHKNSNQIK 60  
 QY 61 IIGNGSFLTKGPSKLNDRADSRSLMDGNFPLIKLIKIEDSDTYICEVEDQKEEYD 120  
 DB 61 IIGKGSFLTKGSSKLRDRIDRSKSLMDGCFSMIKNLIKIEDSEYICEVENKKEEVEI 120  
 QY 121 LVFGLTANSDTHLQGSITLTLSPGSSPSVQCSPPGKXNIGCKTISVQLQDSG 180  
 DB 121 LVFGLTANSDTHLQGSITLTLSPGSSPSVQCSPPGKXNIGCKTISVQLQDSG 180  
 QY 181 TWTCYTLQNKQKVEFDIVLAFQKASSIVYKKEGEQVFSPLAFYTEKLTGSGELMW 240  
 DB 181 TWTCYSDQNTVEFDIDVLAFAQASSTVYKKEGEQVFSPLAFYTEKLTGSGELMW 240  
 QY 241 QERASSSSKSWITFDLKNKESVVKRYTQDPKLQMGKPLPLHLTLPOLPQYAGSGNLTLA 300  
 DB 241 QERASSSSKSWITFDLKNKESVVKRYTQDPKLQMGKPLPLHLTLPOLPQYAGSGNLTLA 300  
 QY 301 LEAKTKLHOENVLVYMRATOLQKNTCEVMPTSPKMLSLKKEKAKVSRERPDVWY 360  
 DB 301 LEAKTKLHOENVLVYMRATOLQKNTCEVMPTSPKMLSLKKEKAKVSRERPDVWY 360  
 QY 361 LNPEAGMOCLSDSGVLLNESIKVLPWSTPVP 394  
 DB 361 LNPEAGMOCLSDSGVLLNESIKVLPWSTPVP 394  
 RESULT 8  
 ID CD4\_ERYPA STANDARD; PRT: 397 AA.  
 AC 008339;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Leu-3)  
 DE (Fragment).  
 GN CD4.  
 OS Erythrocybus patas (Red guenon) (Husar).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;  
 OC Cercopithecoidea; Erythrocybus.  
 NCBI\_Taxid=9538;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX TISSUE=Blood;  
 RX MEDLINE=93049640; Pubmed=1425921;  
 RA Fomsgaard A., Hirsch V.M., Johnson P.R.;  
 RT "Cloning and sequences of primate CD4 molecules: diversity of the  
 RT cellular receptor for simian immunodeficiency virus/human  
 RT immunodeficiency virus.";  
 RL Eur. J. Immunol. 22:2973-2981(1992).  
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell  
 CC receptor interaction. May regulate T-cell activation.



```

CC -1 SUBUNIT: Associates with p56-lck (By similarity).
CC -1 SUBCELLULAR LOCATION: Type I membrane protein.
CC -1 SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1 SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X73324; CAAS1750.1; -.
DR HSSP: P01730; IWI0.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:coreceptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . . ; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007165; P:transmembrane receptor protein tyrosine kin. . . ; ISS.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; IG_1-like.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Lipoprotein; Palmitate.
FT NON_TER 1 1
FT DOMAIN 370 391 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 392 397 POTENTIAL.
FT DOMAIN 392 397 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 392 397 IG-LIKE V-TYPE.
FT DOMAIN 392 397 IG-LIKE C2-TYPE 1.
FT DOMAIN 392 397 IG-LIKE C2-TYPE 2.
FT DOMAIN 392 397 IG-LIKE C2-TYPE 3.
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (BY SIMILARITY).
FT DISULFID 14 82 BY SIMILARITY.
FT DISULFID 128 157 BY SIMILARITY.
FT DISULFID 301 343 BY SIMILARITY.
FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
FT NON_TER 397 397
SQ SEQUENCE 397 AA; 44081 MW; 67887397A6B7EA4F CRC64;
-----
Query Match 50.3%; Score 1717; DB 1; Length 397;
Beet Local Similarity 89.9%; Pred. No. 2.7e-102;
Matches 330; Conservative 15; Mismatches 22; Indels 0; Gaps 0;
-----
QY 28 VLKGGQGVETLTCAQSKSIQFHKNSNOIKINGSGSFLTKGPSKLANDRAASRLM 87
DB 1 VLKGGQGVETLTCAQSKSIQFHKNSNOIKINGSGSFLTKGPSKLANDRAASRLM 60
QY 88 DQGNFLLIKNLIKIEDSDTYICEVEDQEEVQLVFGLTANSDFHLLQGSLTLTLBSP 147
DB 61 DQGNFLLIKNLIKIEDSDTYICEVEDQEEVQLVFGLTANSDFHLLQGSLTLTLBSP 120
QY 148 GSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTTCIVLONOKKVEKIDIVLAFAKA 207
DB 121 GSSPSVQCRSPRGKNIQGGKTLVSQLELDQSGTTCIVLONOKKVEKIDIVLAFAKA 180
QY 208 SSIYKKEGEQVEFSPLAFVTEKLTGSGELMWOERASSKSNITPDLKKKEVSVKVT 267
DB 181 SSIYKKEGEQVEFSPLAFVTEKLTGSGELMWOERASSKSNITPDLKKKEVSVKVT 240
QY 268 QDPKIQMGKPLPLHLTPQALPOYAGSNTLTALFAKTKLHQBENLVVMBATOLQKNT 327
DB 241 QDPKIQMGKPLPLHLTPQALPOYAGSNTLTALFAKTKLHQBENLVVMBATOLQKNT 300

```

```

QY 328 CEVWGPTSPKMLSLKLENKEAKVSKREKPVWVLPAGMOCLLSDSGVLTESNIKVL 387
DB 301 CEVWGPTSPKMLSLKLENKEATISKQAKAVWVLPBEGMOCLLSDSGVLTESNIKVL 360
QY 388 PTWSTPV 394
DB 361 PTWSTPV 367
-----
RESULT 9
CD4_CERTO STANDARD; PRT; 397 AA.
ID CD4_CERTO
AC 008336;
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 (T-cell surface antigen T4/Lew-3)
DE (Fragment).
GN CD4.
OS Cercopithecus torquatus atys (Red-crowned mangabey) (Sooty mangabey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecinae; Cercopithecus.
OX NCBI_TaxID=9531;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Blood;
RX MEDLINE=93049640; PubMed=1425921;
RA Fomsgaard A., Hirsch V.M., Johnson P.R.;
RT "Cloning and sequences of primate CD4 molecules: diversity of the
RT cellular receptor for simian immunodeficiency virus/human
RT immunodeficiency virus";
RL Eur. J. Immunol. 22:2973-2981(1992).
CC -1 FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1 SUBUNIT: Associates with p56-lck (By similarity).
CC -1 SUBCELLULAR LOCATION: Type I membrane protein.
CC -1 SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1 SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL: X73328; CAAS1754.1; -.
DR HSSP: P01730; IWI0.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:coreceptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . . ; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . . ; ISS.
DR InterPro: IPR000973; CD4_TcAg.
DR InterPro: IPR007110; IG_1-like.
DR InterPro: IPR003596; IG_V.
DR Pfam: PF00047; Ig_2.
DR PRINTS: PR00692; CD4TCANTIGEN.
DR SMART: SM00406; IGV_1.
DR PROSITE: PS50835; IG_LIKE_1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Lipoprotein; Palmitate.
FT NON_TER 1 1
FT DOMAIN 370 391 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 392 397 POTENTIAL.
FT DOMAIN 392 397 CYTOPLASMIC (POTENTIAL).

```

```

FT DOMAIN <1 98 IG-LIKE V-TYPE.
FT DOMAIN 99 176 IG-LIKE C2-TYPE 1.
FT DOMAIN 177 290 IG-LIKE C2-TYPE 2.
FT DOMAIN 291 347 IG-LIKE C2-TYPE 3.
FT CARBOHYD 15 15 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 269 269 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 298 298 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT DISULFID 14 82 BY SIMILARITY.
FT DISULFID 128 157 BY SIMILARITY.
FT DISULFID 301 343 BY SIMILARITY.
FT LIPID 392 392 S-palmitoyl cysteine (By similarity).
FT LIPID 395 395 S-palmitoyl cysteine (By similarity).
FT VARIANT 20 20 MISSING.
FT VARIANT 43 43 T -> I.
FT VARIANT 86 86 N -> D.
FT VARIANT 96 96 F -> L.
FT VARIANT 173 173 V -> M.
FT VARIANT 316 316 R -> K.
FT NON TER 397 397
SQ SEQUENCE 397 AA; 43926 MW; 8660B636D2DB38A7 CRC64;

Query Match 50.0%; Score 1707; DB 1; Length 397;
Best Local Similarity 89.6%; Pred. No. 1.2e-101;
Matches 329; Conservative 18; Mismatches 20; Indels 0; Gaps 0;

QY 28 VVLGKGDVLTCTASQKSIQFHMKNSNOIKILNQSGFLTKGSKLNDRADSRSLM 87
DB 1 VVLGKGDVLTCTASQKSIQFHMKNSNOIKILNQSGFLTKGSKLNDRADSRSLM 60
QY 88 DQGNFPLIINKLKIEDSDTYICEVEDKEEVOLLVFGLTANSPTHLLOQSLTTLTSP 147
DB 61 DQGNFPLIINKLKIEDSDTYICEVEDKEEVOLLVFGLTANSPTHLLOQSLTTLTSP 120
QY 148 GSSPSVQCRSPRGKNIQSGKTLVSQLELQDSGTWTCTVLQNGKVEFKIDIVLAFQKA 207
DB 121 GSSPSVQCRSPRGKNIQSGKTLVSQLELQDSGTWTCTVLQNGKVEFKIDIVLAFQKA 180
QY 208 SSIYKKEGROVESFPLATVEKLTJSSGELMWAQEAASSSKSWITFDLNKEVSVKRT 267
DB 181 SSIYKKEGROVESFPLATVEKLTJSSGELMWAQEAASSSKSWITFDLNKEVSVKRT 240
QY 268 QDPPLQMGKPLPHLTLPOALPOYAGSGLTLALFAKTGKLOHEVNLVWRATQLOKNT 327
DB 241 QDPPLQMGKPLPHLTLPOALPOYAGSGLTLALFAKTGKLOHEVNLVWRATQLOKNT 300
QY 328 CEVWGPTSPKLMSTLKNKEAKVSKREKPYWVLPBAGMOCILSDSGVLLSENIKVL 397
DB 301 CEVWGPTSPKLMSTLKNKEAKVSKREKPYWVLPBAGMOCILSDSGVLLSENIKVL 360
QY 388 PTWSTPV 394
DB 361 PTWSTPV 367

RESULT 10
ID-CD4_SAIISC STANDARD; PRT; 457 AA.
AC Q29037;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3).
GN CD4.
OS Saimiri sciureus (Common squirrel monkey).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Platyrrhini; Cebidae; Saimiri.
OX NCBI_TaxID=9521;
RN [1]
RP SEQUENCE FROM N.A.
RA Tacsun M., Hashimoto O.;
RL Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.
-1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell

```

```

CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D86588; BAA1313.1; -.
CC HSSP: P01730; IMR.
CC DR GO: GO:0042101; C: T-cell receptor complex; ISS.
CC DR GO: GO:0015026; P: coreceptor activity; ISS.
CC DR GO: GO:0042289; P: MHC class II protein binding; ISS.
CC DR GO: GO:0006955; P: immune response; ISS.
CC DR GO: GO:0045086; P: positive regulation of interleukin-2 biosyn. .; ISS.
CC DR GO: GO:0030217; P: T-cell differentiation; ISS.
CC DR GO: GO:0045058; P: T-cell selection; ISS.
CC DR GO: GO:0071659; P: transmembrane receptor protein tyrosine kin. .; ISS.
CC DR InterPro: IPR000973; CD4-TCRG.
CC DR InterPro: IPR007110; IG-like.
CC DR InterPro: IPR003596; IG_v.
CC DR Pfam: PF00047; Ig_2.
CC DR PRINTS: PR00692; CD4TCANTIGEN.
CC DR SMART: SM00406; IGv_1.
CC DR PROSITE: PS00835; IG-LIKE; 1.
CC KM Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
CC Immune response; Repeat; Signal; Lipoprotein; Palmitate.
CC BY SIMILARITY.
CC T-CELL SURFACE GLYCOPROTEIN CD4.
CC EXTRACELLULAR (POTENTIAL).
CC POTENTIAL.
CC CYTOPLASMIC (POTENTIAL).
CC IG-LIKE V-TYPE.
CC IG-LIKE C2-TYPE 1.
CC IG-LIKE C2-TYPE 2.
CC IG-LIKE C2-TYPE 3.
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC N-LINKED (GLCNAC. . .) (POTENTIAL).
CC BY SIMILARITY.
CC BY SIMILARITY.
CC S-palmitoyl cysteine (By similarity).
CC S-palmitoyl cysteine (By similarity).
CC SEQUENCE 457 AA; 50871 MW; 57EBD6344005A015 CRC64;

Query Match 46.1%; Score 1572.5; DB 1; Length 457;
Best Local Similarity 78.2%; Pred. No. 5.2e-93;
Matches 308; Conservative 31; Mismatches 54; Indels 1; Gaps 1;

QY 1 MNRGVFPHLLVLTQALLPATQGNKVLGKGGDVELTCTASQKSIQFHMKNSNOIK 60
DB 1 MNRGVFPHLLVLTQALLPATQGNKVLGKGGDVELTCTASQKSIQFHMKNSNOIK 60
QY 61 ILGNQSPFLTKGSKLNDRADSRSLMDQGNFPLIKNKIEDSDTYICEVEDKEEVOL 120
DB 61 ILGNQSPFLTKGSKLNDRADSRSLMDQGNFPLIKNKIEDSDTYICEVEDKEEVOL 120
QY 61 ILGVQNYFVTRGOSKLTDRIDSKSSWMDGSPFLIKDARIEDSEYICEVESKKEVEL 120
DB 61 ILGVQNYFVTRGOSKLTDRIDSKSSWMDGSPFLIKDARIEDSEYICEVESKKEVEL 120
QY 121 LVFGLTANSPTHLLOQSLTTLTLESPGSSPSVQCRSPRGKNIQSGKTLVSQLELQDSG 180
DB 121 LVFGLTANSPTHLLOQSLTTLTLESPGSSPSVQCRSPRGKNIQSGKTLVSQLELQDSG 180
QY 121 QVFGTLRANPDTHLLOQSLTTLTLESPGSSPSVECTSPGKRIRGRKTLVSQLGIPDSG 180
DB 121 QVFGTLRANPDTHLLOQSLTTLTLESPGSSPSVECTSPGKRIRGRKTLVSQLGIPDSG 180
QY 181 TWTCYVLQNGKVEFKIDIVLAFQKASSIVYKKEGROVESFPLATVEKLTJSSGELM 240
DB 181 TWTCYVLQNGKVEFKIDIVLAFQKASSIVYKKEGROVESFPLATVEKLTJSSGELM 240
QY 181 TWKCTVFOHLELV-FEINIVLVAFAQASSIVYKKEGROVESFPLATVEKLTJSSGELM 239
DB 241 QABRASSKSWITFDLNKEVSVKRTVTPDKMGKPLPHLTLPOALPOYAGSGLTLA 300

```

```

Db 240 QAEKSSSSKSWTFNLTKEOVYVVKLVTDPKLRMGKKPLHLTLTAQALPOVAGSNTFLTA 299
Qy 301 LEAKTGKHOEYNLVVMBATOLQKRLTCEVWGPTSPKMLSLKLENKAKYSKREKPYWV 360
Db 300 LKKGTKGKHOEYNLVVMBATOLQKRLTCEVWGPTSPKMLSLKLENKAKYSKREKAVWV 359
Qy 361 LNPEAGWMOCLLSDSGQVLLSEINIKVLPWTSTPV 394
Db 360 LNPEAGWMOCLLSDSGQVLLSEKFEALPTRSPPV 393

RESULT 11
GCL HUMAN STANDARD; PRT; 330 AA.
ID _GCL_HUMAN
AC P01857;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig gamma-1 chain C region.
GN IGHG1.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=82274238; PubMed=6287432;
RA Ellison J.W., Berson B.J., Hood L.E.;
RT "The nucleotide sequence of a human immunoglobulin C gamma1 gene.";
RL Nucleic Acids Res. 10:4071-4079(1982).
RN [2]
RP SEQUENCE OF 1-135 (MYELOMA PROTEIN EU).
RX MEDLINE=71064024; PubMed=5489771;
RA Cunningham B.A., Rutishauser U., Gall W.E., Gottlieb P.D.,
RA Maxdel M.J., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. VII. Amino
RL Biochemistry 9:3161-3170(1970).
RN [3]
RP SEQUENCE OF 136-329 (EU).
RX MEDLINE=71064025; PubMed=5530842;
RA Rutishauser U., Cunningham B.A., Bennett C., Konigsberg W.H.,
RA Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. 8. Amino
RL Biochemistry 9:3171-3181(1970).
RN [4]
RP SEQUENCE (MYELOMA PROTEIN NIE).
RX MEDLINE=77070269; PubMed=826475;
RA Ponstingl H., Hilschmann N.;
RT "The rule of antibody structure. The primary structure of a
RL monoclonal IgG1 immunoglobulin (myeloma protein Nie). III. The
RT chymotryptic peptides of the H-chain, alignment of the tryptic
RL peptides and discussion of the complete structure.";
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1571-1604(1976).
RN [5]
RP SEQUENCE (MYELOMA PROTEIN KOL), AND DISULFIDE BONDS.
RX MEDLINE=83289131; PubMed=6884994;
RA Schmidt W.E., Jung H.-D., Palm W., Hilschmann N.;
RT "Three-dimensional structure determination of antibodies. Primary
RT structure of crystallized monoclonal immunoglobulin IgG1 KOL, I.",
RL Hoppe-Seyler's Z. Physiol. Chem. 364:713-747(1983).
RN [6]
RP DISULFIDE BONDS.
RX MEDLINE=71064027; PubMed=4923144;
RA Gall W.E., Edelman G.M.;
RT "The covalent structure of a human gamma G-immunoglobulin. X.
RL Interchain disulfide bonds.";
RL Biochemistry 9:3188-3196(1970).
RN [7]
RP DISULFIDE BONDS.
RX MEDLINE=77070267; PubMed=1002129;

```

```

RA Dreker L., Schwarz J., Reichel W., Hilschmann N.;
RT "Rule of antibody structure. The primary structure of a monoclonal
RT IgG1 immunoglobulin (myeloma protein Nie). I: Purification and
RT characterization of the protein, the L- and H-chains, the
RL cytochrome brome cleavage products, and the disulfide bridges.";
RL Hoppe-Seyler's Z. Physiol. Chem. 357:1515-1540(1976).
RN [8]
RP X-RAY CRYSTALLOGRAPHY (2.9 ANGSTROMS).
RX MEDLINE=81208100; PubMed=7236608;
RA Delsehofer J.;
RT "Crystallographic refinement and atomic models of a human Fc fragment
RT and its complex with fragment B of protein A from Staphylococcus
RT aureus at 2.9- and 2.8-A resolution.";
RL Biochemistry 20:2361-2370(1981).
CC -1- MISCELLANEOUS: Nie has the GIM(17) allotypic marker, 97-K, and the
CC GIM(1) markers, 239-D and 241-L. KOL and EU sequences have the
CC GIM(3) marker and the GIM (non-1) markers.
CC -1- MISCELLANEOUS: Nie also differs in the amidation states of
CC 35, 116, 198, 269 and 272.
CC -1- MISCELLANEOUS: EU also differs in the amidation states of residues
CC 155, 166, 177, 195, 198, 269, and 272 and in the order of residues
CC 268-272.
CC -1- MISCELLANEOUS: KOL also differs in the amidation states of
CC residues 198, 267 and 272.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00228; AAC82527.1; ALT_INIT.
DR PIR; A93433; GHNU.
DR PDB; 1FC1; 15-JUL-92.
DR PDB; 1FC2; 15-JUL-92.
DR PDB; 1AJ7; 12-NOV-97.
DR PDB; 1D57; 09-FEB-00.
DR PDB; 1D51; 09-FEB-00.
DR PDB; 1D6V; 04-OCT-00.
DR PDB; 1DN2; 17-MAY-00.
DR PDB; 1E4K; 06-JUN-01.
DR PDB; 1ECC; 20-JUL-95.
DR PDB; 1HZH; 12-JUN-02.
DR PDB; 1I7Z; 08-AUG-01.
DR PDB; 1I1S; 16-MAY-01.
DR PDB; 1I1X; 16-MAY-01.
DR PDB; 1L6X; 10-APR-02.
DR PDB; 2RCS; 12-NOV-97.
DR Genew; HGNC:5525; IGHG1.
DR MIM; 147100; -.
DR GO; GO:0005624; C:membrane fraction; NAS.
DR GO; GO:0003823; F:antigen binding; TAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR Pfam; PF00047; Ig_3.
DR SMART; SM00407; IGH1; 2.
DR PROSITE; PSS0835; IG_LIKE; 3.
DR PROSITE; PSS0290; IG_MHC; 2.
DR Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW 3D-structure.
FT NON TER 1 1
FT DOMAIN 98 98 CH1.
FT DOMAIN 99 110 HINGE.
FT DOMAIN 111 223 CH2.
FT DOMAIN 224 330 CH3.
FT DISULFID 27 83
FT DISULFID 103 103 INTERCHAIN (WITH LIGHT CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH HEAVY CHAIN).

```

```
FT DISULFID 144 204
FT DISULFID 250 308
FT CARBOHYD 180 180
FT MOD RES 330 330
FT VARIANT 97 97
FT VARIANT 239 239
FT VARIANT 241 241
FT STRAND 122 126
FT HELIX 130 134
FT TURN 136 137
FT STRAND 141 147
FT STRAND 157 162
FT TURN 163 164
FT STRAND 165 166
FT TURN 168 171
FT STRAND 176 179
FT TURN 180 181
FT STRAND 182 190
FT HELIX 193 197
FT TURN 198 199
FT STRAND 202 207
FT TURN 209 210
FT STRAND 215 219
FT STRAND 227 227
FT STRAND 230 234
FT HELIX 238 242
FT STRAND 245 256
FT STRAND 260 265
FT STRAND 270 270
FT STRAND 274 276
FT TURN 280 281
FT STRAND 283 284
FT STRAND 287 296
FT HELIX 297 301
FT TURN 302 303
FT TURN 305 312
FT TURN 313 314
FT TURN 316 317
FT STRAND 320 325
SQ SEQUENCE 330 AA; 36106 MM; 3770EHL06C2FA33D CRC64;

Query Match 35.5%; Score 1211.5; DB 1; Length 330;
Best Local Similarity 97.8%; Pred. No. 3.3e-70;
Matches 226; Conservative 0; Mismatches 0; Indels 5; Gaps 1;

QY 400 EPKSCDKTHRC-----PELLGSPVFLPPKPKDTLMISRTPEVTCVVVSHEDPEVKF 454
DB 99 EPKSCDKTHRCPPCPAPPELLGGPSVFLFPPPKDTLMISRTPEVTCVVVSHEDPEVKF 158
QY 455 NMVYDGEVHNNAKTPREBOYNSTYRVSVLTVLHODMLNGKEYCKVSNKALPAPIEKT 514
DB 159 NMVYDGEVHNNAKTPREBOYNSTYRVSVLTVLHODMLNGKEYCKVSNKALPAPIEKT 218
QY 515 ISRAKGPREPQVYTLPPSDELTKNOVSLTCLVKGFPEDIAVEMESNQPENNYKTP 574
DB 219 ISRAKGPREPQVYTLPPSDELTKNOVSLTCLVKGFPEDIAVEMESNQPENNYKTP 278
QY 575 PVLDSDSFLLSKLTVDKSRMOGNVSCSVMEALAHNYTKSLSPG 625
DB 279 PVLDSDSFLLSKLTVDKSRMOGNVSCSVMEALAHNYTKSLSPG 329

RESULT 12
CD4_RABIT STANDARD; PRT; 459 AA.
AC P46630;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
```

```
DE T4/Leu-3).
GN CD4.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid:9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92390370; PubMed=1518821;
RA Hague B.F., Sawaadikosol S., Brown T.J., Lee K., Recker D.P.,
RA Kindt T.J.;
RT "CD4 and its role in infection of rabbit cell lines by human
RT immunodeficiency virus type 1."
RL Proc. Natl. Acad. Sci. U.S.A. 89:7963-7967(1992).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M92840; AAA31198.1; -.
CC PIR; A46254; A46254.
CC HSSP; P01730; 1MBR.
DR GO; GO:0042101; C: T-cell receptor complex; ISS.
DR GO; GO:0015026; F: coreceptor activity; ISS.
DR GO; GO:0042289; F: MHC class II protein binding; ISS.
DR GO; GO:006955; P: immune response; ISS.
DR GO; GO:0045086; P: positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO; GO:0030217; P: T-cell differentiation; ISS.
DR GO; GO:0045058; P: T-cell selection; ISS.
DR GO; GO:0007169; P: transmembrane receptor protein tyrosine kin. . .; ISS.
DR InterPro; IPR000973; CD4_TCMG.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003596; Ig_V.
DR Pfam; PF00047; Ig_2.
DR PRINTS; PR00692; CD4TCANTIGEN.
DR SMART; SM00406; IGV_1.
DR PROSITE; PS50835; IG_LIKE; 1.
KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
KW Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT SIGNAL 1 25
FT CHAIN 26 459
FT DOMAIN 26 396
FT TRANSMEM 397 419
FT DOMAIN 420 459
FT DOMAIN 26 129
FT DOMAIN 130 208
FT DOMAIN 209 318
FT DOMAIN 319 374
FT CARBOHYD 299 299
FT DISULFID 41 113
FT DISULFID 329 370
FT LIPID 420 420
FT LIPID 423 423
SQ SEQUENCE 459 AA; 50866 MM; B323311CBDA0013D CRC64;

Query Match 33.7%; Score 1150.5; DB 1; Length 459;
Best Local Similarity 57.2%; Pred. No. 3.9e-66;
Matches 241; Conservative 65; Mismatches 90; Indels 25; Gaps 6;

QY 1 MNRGVPFRLLVLTQLALLPAATQGNKVVYLGKKGDVLELTCTASQKKSQIFHKNNSNQIK 60
DB 1 MNRRIYFQGLLVLPALLPAATWGTQVVRGKAGALVELPCQSSQKNSVFNNWGHANQVK 60
```



```

FT  VARIANT      279      279      /FTId=VAR_003895.
FT  SEQUENCE      290 AA; 32331 MW; E69CBG95705B2F46 CRC64;
SQ
Query Match      33.5%; Score 1143; DB 1; Length 290;
Best Local Similarity 87.9%; Pred. No. 6, 4e-66;
Matches 211; Conservative 13; Mismatches 10; Indels 6; Gaps 2;

QY  392 TVPCCP-APPPKSCDKTHTC-----PELLGSPVFLPPPKKDTLMISRTPEVTCVVVDY 445
DB  50 TPSPCCRCPPKSCDPPPPPCPCPAPLLGSPVFLPPPKKDTLMISRTPEVTCVVVDY 109
QY  446 SHEDPEVKFWYVDGVVHNAKTKRPREQVNSTYRVAVSVTLVHOMLNKEVKCKVSNK 505
DB  110 SHEDPEVQFKWYVDGVVHNAKTKRPRQGNSTFRVAVSVTLVHOMLNDEKVKCKVSNK 169
QY  506 ALPAPIEKTSKAGQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMESNQ 565
DB  170 ALPAPIEKTSKTKGQPREPQVYTLPPSRDEMTKNQVSLTCLVKGFYPSDIAVEMESSGQ 229
QY  566 PENNYKTPPVLDSDGSFPLYSKLTVDKSRWQQGNVFCGVMHEDALHNYTKQSLSPG 625
DB  230 PENNYKTPPVLDSDGSFPLYSKLTVDKSRWQQGNVFCGVMHEDALHNYTKQSLSPG 289

RESULT 14
CD4 CANFA
ID CD4 CANFA STANDARD; PRT; 463 AA.
AC P33705;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
   T4/Lew-3).
GN CD4.
OS Canis familiaris (Dog).
OC Eukaryote; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN 1)
RP SEQUENCE OF 13-463 FROM N.A.
RC STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=93192324; PubMed=791632;
RA Milde K.F., Conner G.E., Minz D.H., Alejandro R.;
RT "Primary structure of the canine CD4 antigen.";
RL Biochim. Acta 1172:315-318(1993).
RN 12)
RP SEQUENCE FROM N.A.
RC STRAIN=Beagle; TISSUE=Thymus;
RX MEDLINE=94378217; PubMed=8091416;
RA Gorman S.D., Fretwin M.R., Cobbold S.P., Waldmann H.;
RT "Isolation and expression of cDNA encoding the canine CD4 and CD8
   alpha antigens.";
RL Tissue Antigens 43:184-188(1994).
CC -!- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
   receptor interaction. May regulate T-cell activation.
CC -!- SUBUNIT: Associates with p56-lck (by similarity).
CC -!- SUBCELLULAR LOCATION: Type I membrane protein.
CC -!- TISSUE SPECIFICITY: Expressed in macrophages and a subset of
   T lymphocytes.
CC -!- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.
CC -!- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC EMBL; L06130; AAB02295.1; -

```

```

DR  EMBL; X68565; -; NOT_ANNOTATED_CDS.
DR  HSSP; P01730; 1MR.
DR  GO; GO:0042101; C1T-cell receptor complex; ISS.
DR  GO; GO:0015026; F:coreceptor activity; ISS.
DR  GO; GO:0042289; F:MHC class II protein binding; ISS.
DR  GO; GO:0006955; P:immune response; ISS.
DR  GO; GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR  GO; GO:0030217; P:T-cell differentiation; ISS.
DR  GO; GO:0045058; P:T-cell selection; ISS.
DR  GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.
DR  InterPro; IPR000973; CD4_TcRg.
DR  InterPro; IPR007110; Ig-like.
DR  InterPro; IPR003596; Ig_V.
DR  Pfam; PF00047; Ig_3.
DR  PRINTS; PR00692; CD4TCANTIGEN.
DR  SMART; SM00406; IGV; 1.
DR  PROSITE; PS00835; IG_LIKE; 1.
DR  KMW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;
   Immune response; Repeat; Signal; Lipoprotein; Palmitate.
FT  SIGNAL 1 24
FT  CHAIN 1 463
FT  DOMAIN 25 401
FT  TRANSMEM 402 423
FT  DOMAIN 424 463
FT  DOMAIN 26 124
FT  DOMAIN 125 211
FT  DOMAIN 212 321
FT  DOMAIN 322 378
FT  DISULFID 41 109
FT  LIPID 424 424
FT  LIPID 427 427
FT  CARBOHYD 123 123
FT  CARBOHYD 168 168
FT  CARBOHYD 176 176
FT  CARBOHYD 324 324
FT  CARBOHYD 329 329
FT  CARBOHYD 389 389
SQ  SEQUENCE 463 AA; 51639 MW; 95805170CB4A833 CRC64;

Query Match      33.2%; Score 1135; DB 1; Length 463;
Best Local Similarity 57.8%; Pred. No. 3, 8e-65;
Matches 233; Conservative 63; Mismatches 89; Indels 18; Gaps 6;

QY  1 NMRGVFRRLLVQLALLPATQGNKYVLGKKGDVVELTCTRSOKKSTOFHKNKNQIK 60
DB  1 MNQEAFFRHLMLQLVMLPAVTPVREVLGKAGDAVELPCQTSQKKNIHFNRDSSMWQ 60
QY  61 ILGNQGSFLLTKGSKLNDRAADRSRLSDPGNPFLLIKNKIEDSDTYICEVDQKEEVQL 120
DB  61 ILGNQGSFMTVGSRLKHKVESKKNLMDQSPFLVKDELVADSGIYFCDT-DKRGQVEL 119
QY  121 LVFGLTAA-----NSDTHLLQGSLLTLTLESPPGSSPSVQCRSPRGKNIQGGKTLVY 171
DB  120 LVFNLTAKWDSCSGSSSNIRLLQGOQLTLTLENPGSSPSVQWKGPKNKSFKGCGNLSL 179
QY  172 SQLELDSDGTWCTVQGNQKVEFKDIYVLAFOKXSIYVKKEGOVESFPLATVEK 231
DB  180 SWEILDGGGTWCTTISQSKTVEFNINVLVLAFOKXSNFYAEGDQVESFPLSPEDEN 239
QY  232 LTSGSGELMWQAEKASSKSMITFDLNKKEVSVKRVTDPLQMGKKLPLHLTPQALPOY 291
DB  240 LV-GEIIRQDAQASSSLMISTLENKLSKMEALAPLKLQKESLPLRFTLPQVLSRY 297
QY  292 AGSGNLTLLAEATGTLHOEVNLVVRATQLOKNLTCEVWGPTSPKLMSLKLENKAVY 351
DB  298 AGSGIITLNL-AK-GTLVQEVNLVVRANSSQNNLTCEVLPGPSPELTLSLNKEQAKV 355
QY  352 SKREKRVVNLNPAQMWQCLISDQVYLLSEINKYLPMTSTVP 394
DB  356 SKQKLVWVVDPEGGTWQCLISDKDKVLLASSLNV-----SSPV 394

```

RESULT 15  
 ID GC2 HUMAN STANDARD; PRT; 326 AA.  
 AC P01859;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Ig gamma-2 chain C region.  
 GN IGHG2.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_Taxid:9606;  
 RN [1]  
 RP SEQUENCE OF 2-326 FROM N.A.  
 RX MEDLINE=82197621; PubMed=6604948;  
 RA Ellison J.W., Hood L.E.;  
 RT "Linkage and sequence homology of two human immunoglobulin gamma  
 heavy chain constant region genes.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:1984-1988(1982).  
 RN [2]  
 RP SEQUENCE OF 88-115 FROM N.A.  
 RC TISSUE=Fetal liver;  
 RX MEDLINE=83001943; PubMed=6811339;  
 RA Takahashi N., Ueda S., Obata M., Nikaido T., Nakai S., Honjo T.;  
 RT "Structure of human immunoglobulin gamma genes: implications for  
 evolution of a gene family.";  
 RL Cell 29:671-679(1982).  
 RN [3]  
 RP SEQUENCE OF 99-177 AND 310-326 FROM N.A.  
 RC TISSUE=Fetal liver;  
 RX MEDLINE=84235992; PubMed=6329676;  
 RA Krawinkel U., Rebhitzer T.H.;  
 RT "Comparison of the hinge-coding segments in human immunoglobulin gamma  
 heavy chain genes and the linkage of the gamma 2 and gamma 4 subclass  
 genes.";  
 RL EMBO J. 1:403-407(1982).  
 RN [4]  
 RP SEQUENCE OF 1-325 (MYELOMA PROTEIN TIL).  
 RX MEDLINE=81007873; PubMed=6774012;  
 RA Wang A.-C., Tung E., Fudenberg H.H.;  
 RT "The primary structure of a human IgG2 heavy chain: genetic,  
 evolutionary, and functional implications.";  
 RL J. Immunol. 125:1048-1054(1980).  
 RN [5]  
 RP SEQUENCE OF 1-85 AND 132-325 (MYELOMA PROTEIN ZIE).  
 RX MEDLINE=80001357; PubMed=113060;  
 RA Connell G.E., Parr D.M., Hofmann T.;  
 RT "The amino acid sequences of the three heavy chain constant region  
 domains of a human IgG2 myeloma protein.";  
 RL Can. J. Biochem. 57:758-767(1979).  
 RN [6]  
 RP SEQUENCE OF 238-275 (ZIE).  
 RX MEDLINE=8011419; PubMed=118920;  
 RA Hofmann T., Parr D.M.;  
 RT "A note of the amino acid sequence of residues 381-391 of human  
 immunoglobulin gamma chains.";  
 RL Mol. Immunol. 16:923-925(1979).  
 RN [7]  
 RP REVISIONS TO 25; 59; 60 AND 264-268 (ZIE).  
 RA Hofmann T., Parr D.M.;  
 RL Submitted (MAR-1980) to the PIR data bank.  
 RN [8]  
 RP SEQUENCE OF 1-121 (DOT).  
 RX MEDLINE=95255298; PubMed=7737190;  
 RA Scoppini M., Bellotti V., Negri A., Merlini G., Garver F., Ferri G.;  
 RT "Characterization of the two unique human anti-flavin monoclonal  
 immunoglobulins.";  
 RL Eur. J. Biochem. 228:886-893(1995).  
 RN [9]  
 RP DISULFIDE BONDS.  
 RX MEDLINE=72033500; PubMed=4940472;  
 RA Milstien C., Frangione B.;

RT "Disulphide bridges of the heavy chain of human immunoglobulin G2.";  
 RL Biochem. J. 121:217-225(1971).  
 RN [10]  
 RP DISULFIDE BONDS.  
 RX Frangione B., Milstien C., Pink J.R.L.;  
 RA "Structural studies of immunoglobulin G.";  
 RT Nature 221:145-148(1969).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements/>  
 CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
 CC -----  
 DR EMBL; J00230; AAB59393.1; -.  
 DR PIR; A93906; G2HU.  
 DR HSSP; P01857; 1FC1.  
 DR Genew; HGNC:5526; IGHG2.  
 DR MIM; 147110; -.  
 DR GO; GO:0005624; C:membrane fraction; NAS.  
 DR GO; GO:0003823; F:antigen binding; TAS.  
 DR GO; GO:0006955; P:immune response; NAS.  
 DR InterPro; IPR007110; IG-like.  
 DR InterPro; IPR003597; IG C1.  
 DR InterPro; IPR003006; IG\_MHC.  
 DR Pfam; PF00047; Ig; 3.  
 DR SMART; SM00407; IGC1; 2.  
 DR PROSITE; PS00835; IG LIKE; 3.  
 DR PROSITE; PS00290; IG\_MHC; 2.  
 KM Immunoglobulin domain; Immunoglobulin C region.  
 FT NON\_TER 1 1  
 FT DOMAIN 1 98  
 FT 99 110 CH1.  
 FT DOMAIN 111 219 HINGE.  
 FT 220 326 CH2.  
 FT DOMAIN 14 14 CH3.  
 FT DISULFID 27 83 INTERCHAIN (WITH A LIGHT CHAIN).  
 FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT DISULFID 103 103 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT DISULFID 140 200  
 FT DISULFID 246 304  
 FT DISULFID 156 156  
 FT SITE 156 326  
 FT MOD\_RES 326 326  
 FT VARIANT 60 60  
 FT CONFLICT 109 109  
 FT SEQ 326 AA; 35884 MW; 83108786C6878CF9C CRC64;  
 SQ  
 Query Match 32.7%; Score 1115.5; DB 1; Length 326;  
 Best Local Similarity 89.2%; Pred. No. 4.2e-64;  
 Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;  
 QY 395 PCPAPBPKSCDKHTCPCLLGGPSVFLPPKPKDTLMISRTPEVTCVAVDVSHPDPEVKF 454  
 DB 108 PCPAPP-----VAGPSVFLPPKPKDTLMISRTPEVTCVAVDVSHPDPEVQF 154  
 QY 455 NMVYDGVVHNNAKTPREEDQNSTFRVSVLTIVHQMVGKEYKCVSNKALPAPLEKT 514  
 DB 155 NMVYDGVVHNNAKTPREEDQNSTFRVSVLTIVHQMVGKEYKCVSNKGLPAPLEKT 214  
 QY 515 ISKAGQPREQVYTLPPSRDELTKQVSLTCLVKGYPSPDIADVEMSGQPENNYKTPP 574  
 DB 215 ISKTKGPREQVYTLPPSRDEMTKQVSLTCLVKGYPSPDIADVEMSGQPENNYKTPP 274  
 QY 575 PVLDSDGSPFLYSKLTVDKSRMOQGNVPSGVMEALHNHYTKSLSLSPG 625  
 DB 275 PVLDSDGSPFLYSKLTVDKSRMOQGNVPSGVMEALHNHYTKSLSLSPG 325

```

RESULT 16
ID GC4_HUMAN STANDARD; PRT; 327 AA.
AC P01861;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-4 chain C region.
GN IGHG4.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=83157104; PubMed=6299662;
RA Ellison J.W., Buxbaum J.N., Hood L.E.;
RT "Nucleotide sequence of a human immunoglobulin C gamma 4 gene.";
RL DNA 1:11-18(1981).
RN [2]
RN SEQUENCE OF 1-30 AND 81-326.
RX MEDLINE=70207560; PubMed=4192699;
RA Pink J.R.L., Buttery S.H., de Vries G.M., Milstein C.;
RT "Human immunoglobulin subclasses. Partial amino acid sequence of the
RT constant region of a gamma 4 chain.";
RL Biochem. J. 117:33-47(1970).
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: K01316; AAB59394.1; ALT_INT.
DR PIR: A90833; GAHU.
DR PDB: 1ADQ; 16-SEP-98.
DR Genew; HGNC:5528; IGHG4.
DR MIM; 147130; -.
DR GO: GO:0005624; C:membrane fraction; NMS.
DR GO: GO:0003823; F:antigen binding; TAS.
DR GO: GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PSS0835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KM Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
FT NON_TER 1 1
FT DOMAIN 1 98 CH1.
FT DOMAIN 99 110 HINGE.
FT DOMAIN 111 220 CH2.
FT DOMAIN 221 327 CH3.
FT DISULFID 14 14 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 83
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 141 201
FT DISULFID 247 305
SQ SEQUENCE 327 AA; 35940 MW; 3EBD811EF208E7A CRC64;
Query Match 32.4%; Score 1105.5; DB 1; Length 327;
Best Local Similarity 89.6%; Pred. No. 1.8e-63;
Matches 207; Conservative 8; Mismatches 7; Indels 9; Gaps 2;
QY 395 PCPAPERKSCDKHTCTCPCLLGGPSVFLPFPKQDTLMISRTPEVTCVVVDSHEDPEVK 454
DB 105 PCP-----SC-----PAPEFLGSPSVFLFPFKQDTLMISRTPEVTCVVVDSHEDPEVK 155

```

```

QY 455 NMVYDGEVHNAAKTPREEOVNSTYRVVSVLTFLTHQDMWINGKRYCKVSNKALPAPIEKT 514
DB 156 NMVYDGEVHNAAKTPREEOVNSTYRVVSVLTFLTHQDMWINGKRYCKVSNKALPAPIEKT 215
QY 515 ISKAGQPREPOVYTLTPPSRDELTKYNQVSLTCLVKGFPSPDIAVEMESNGQPENNYKTP 574
DB 216 ISKAGQPREPOVYTLTPPSRDELTKYNQVSLTCLVKGFPSPDIAVEMESNGQPENNYKTP 275
QY 575 PVLDSGSEFFLYSKLTVDKSRWQGNVPSCSFMHMLNHNHYOKSLSLSPG 625
DB 276 PVLDSGSEFFLYSKLTVDKSRWQGNVPSCSFMHMLNHNHYOKSLSLSLG 326
RESULT 17
ID CD4_RAT STANDARD; PRT; 457 AA.
AC P05540;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen
DE T4/Leu-3) (W3/25 antigen).
GN CD4.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RX MEDLINE=87175535; PubMed=3104900;
RA Clark S.J., Jefferies W.A., Barclay A.N., Gagnon J., Williams A.F.;
RT "Peptide and nucleotide sequences of rat CD4 (W3/25) antigen:
RT evidence for derivation from a structure with four
RT immunoglobulin-related domains.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:1649-1653(1987).
RN [2]
RN X-RAY CRYSTALLOGRAPHY (2.8 ANGSTROMS) OF 210-393.
RX MEDLINE=93262437; PubMed=8493535;
RA Brady R.L., Dodson E.J., Dodson G.G., Lange G., Davis S.J.,
RA Williams A.F., Barclay A.N.;
RT "Crystal structure of domains 3 and 4 of rat CD4: relation to the
RT NH2-terminal domains.";
RL Science 260:979-983(1993).
CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell
CC receptor interaction. May regulate T-cell activation.
CC -1- SUBUNIT: Associates with p56-lck (by similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like C2-type domain.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M15768; AAA40901.1; -.
DR PIR: A27449; A27449.
DR PDB: 1CID; 15-JUL-93.
DR GlycoStatedB; P05540; -.
DR GO: GO:0042101; C:T-cell receptor complex; ISS.
DR GO: GO:0015026; F:coreceptor activity; ISS.
DR GO: GO:0042289; F:MHC class II protein binding; ISS.
DR GO: GO:0006955; P:immune response; ISS.
DR GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.
DR GO: GO:0030217; P:T-cell differentiation; ISS.
DR GO: GO:0045058; P:T-cell selection; ISS.
DR GO: GO:0007169; P:T-cell selection; ISS.
DR InterPro; IPR000973; CD4_TGAG.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003599; Ig.

```



DR Pfam; PF00047; Ig; 2.  
 DR PRINTS; PR00692; CD4TCANTIGEN.  
 DR SMART; SMO0409; IG; 2.  
 DR PROSITE; PS00835; IG\_LIKE; 1.  
 KM Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;  
 KM Immune response; Repeat; Signal; Lipoprotein; Palmitate; 3D-structure.  
 FT SIGNAL 1 27  
 FT CHAIN 28 457  
 FT DOMAIN 28 394  
 FT TRANSMEM 395 417  
 FT DOMAIN 418 457  
 FT DOMAIN 128 127  
 FT DOMAIN 207 316  
 FT DOMAIN 317 374  
 FT CARBOHYD 186 186  
 FT CARBOHYD 297 297  
 FT CARBOHYD 392 392  
 FT DISULFID 43 111  
 FT DISULFID 158 187  
 FT DISULFID 328 370  
 FT LIPID 418 418  
 FT LIPID 421 421  
 FT STRAND 213 217  
 FT TURN 218 219  
 FT STRAND 222 225  
 FT STRAND 235 243  
 FT STRAND 252 258  
 FT TURN 259 260  
 FT STRAND 261 265  
 FT STRAND 274 275  
 FT STRAND 278 278  
 FT TURN 282 280  
 FT STRAND 282 285  
 FT HELIX 290 292  
 FT STRAND 294 301  
 FT STRAND 306 319  
 FT STRAND 325 331  
 FT STRAND 338 344  
 FT TURN 345 346  
 FT STRAND 349 353  
 FT STRAND 357 361  
 FT STRAND 367 374  
 FT TURN 375 376  
 FT STRAND 377 385  
 SQ SEQUENCE 457 AA; 51437 MW; 477BE157D30954C1 CRC64;  
 Query Match 29.3%; Score 1000.5; DB 1; Length 457;  
 Best Local Similarity 50.1%; Pred. No. 1.4e-56;  
 Matches 213; Conservative 64; Mismatches 127; Indels 21; Gaps 6;  
 QY 1 MNRGVPRHL--LVVLQALALPAATQGNKVVYLGKKGDVLTCTASQKKSIOFHMKNQ 58  
 DB 1 MCRGFSFHLPLLLQLSKLVVQGTVVLGKGGSAELPCESTRSRASAFMKSSDQ 60  
 QY 59 IKIINGQSFLLTKGSKLMDRADRSRLMDQGNPFLIKLKIIBSDTYICEVEDQKEV 118  
 DB 61 KTIIGYKKLKILKSLLEYSPDSRKNMWERGSPFLINKLRMEDSQTYVELENKKEEV 120  
 QY 119 QLVVGLTANSDTHLLQGSITLTLES--PGSSPSVQCRSPRGKNIQGGKTLVSQLEIQ 177  
 DB 121 ELWVFRVTFNPCTRLQGSITLTILDSNPKVSDPIECKHSSNIVKSKAFSTHSRLIQ 180  
 QY 178 DSGWTCTVLQNOKKVEFKIDI VVLAFOKASSIYKKEGEVPSFLAFTVEKLTGSGE 237  
 DB 181 DSGIMNCTVLTNOKKGHSDMKLSVYGFASTSITAVKSGEASERFPLNLGEESI--QGE 238  
 QY 238 LMQAERASSSKSWTTPDLKXKVEVKVQDPKQKMKLPHLHTLPQALPOVAGSGNL 297  
 DB 239 LRMKAEKRPSSQSMTFSLKQKQVSVQKSTSNPKFQLESTLPLTQIPQVSLQFAGSGNL 298  
 QY 298 TLAEAKTGKHOEVNLVVMRATOLQKN-LTCEVWGPRSPKMLSLKLENKAEKSKREK 356

DB 299 TLALD--RGILYQEVNLVVMKVTPDSDNTLTCEVWGPTSPKMRLLIKQENQEARVSRQEK 356  
 QY 357 PWTALNEAGWMOCLSDSQVLLSENIKILPTMTSTVPCAPRPKSCDKHTCELLGG 416  
 DB 357 VTQVQAEAGVMOCLSEGEVKKMSKIQVL-----SKGLNQTMFLAVLGS 403  
 QY 417 PSVFL 421  
 DB 404 AFSL 408  
 RESULT 18  
 ID CD4 MOUSE STANDARD; PRT; 457 AA.  
 AC P06332;  
 DT 01-JAN-1988 (Rel. 06, Created)  
 DT 01-JAN-1988 (Rel. 06, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE T-cell surface glycoprotein CD4 precursor (T-cell surface antigen T4/Leu-3) (T-cell differentiation antigen L3T4).  
 GN CD4.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_Taxid=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87018845; PubMed=3094146;  
 RA Tourville B., Gorman S.D., Field E.H., Hunkapiller T., Parnes J.R.;  
 RT "Isolation and sequence of L3T4 complementary DNA clones: expression in T cells and brain";  
 RL Science 234:610-614(1986).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=87115821; PubMed=3027575;  
 RA Littman D.R., Gettner S.N.;  
 RT "Unusual intron in the immunoglobulin domain of the newly isolated murine Cd4 (L3T4) gene";  
 RL Nature 325:453-455(1987).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=88152875; PubMed=3326818;  
 RA Parnes J.R., Hunkapiller T.;  
 RT "L3T4 and the immunoglobulin gene superfamily: new relationships between the immune system and the nervous system";  
 RL Immunol. Rev. 100:109-127(1987).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC TISSUE=Brain;  
 RX MEDLINE=88041159; PubMed=2823269;  
 RA Gorman S.D., Tourville B., Parnes J.R.;  
 RT "Structure of the mouse gene encoding CD4 and an unusual transcript in brain";  
 RL Proc. Natl. Acad. Sci. U.S.A. 84:7644-7648(1987).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=98112780; PubMed=9445485;  
 RA Aneari-Lari M.A., Oeltgen J.C., Schwartz S., Zhang Z., Muzny D.M., Lu J., Gorrell J.H., Chinault A.C., Belmont J.W., Miller W., Gibbs R.A.;  
 RT "Comparative sequence analysis of a gene-rich cluster at human chromosome 12p13 and its syntenic region in mouse chromosome 6";  
 RL Genome Res. 8:29-40(1998).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=C57BL/6J; TISSUE=Mammary gland;  
 RX MEDLINE=22386257; PubMed=12477932;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G., Kleeschul S.F., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D., Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heide F., Datchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,

RA Strapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carinici P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Phley J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalls D.E.,  
 RA Schmech A., Schein J.E., Jones S.J.M., Marra M.A.,  
 RT "Generation and initial analysis of more than 15,000 full-length  
 human and mouse cDNA sequences.",  
 RT Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [7]  
 RP SEQUENCE OF 27-43.  
 RX MEDLINE=6166694; PubMed=3082751;  
 RA Classon B.J., Tsagaratos J., Kitzbaum L., Maddox J., McKay C.R.,  
 RA Brandon M., McKenzie I.F.C., Walker I.D.,  
 RT "The L314 antigen in mouse and the sheep equivalent are  
 RT Immunoglobulin-like.",  
 RL Immunogenetics 23:129-132(1986).  
 RN [8]  
 RP DISULFIDE BONDS.  
 RX MEDLINE=6233454; PubMed=3086886.  
 RA Classon B.J., Tsagaratos J., McKenzie I.F.C., Walker I.D.,  
 RT "Partial primary structure of the T4 antigens of mouse and sheep:  
 RT assignment of interchain disulfide bonds.",  
 RL Proc. Natl. Acad. Sci. U.S.A. 83:4499-4503(1986).  
 CC -1- FUNCTION: Accessory protein for MHC class-II antigen/T-cell  
 CC receptor interaction. May regulate T-cell activation.  
 CC -1- SUBUNIT: Associates with p56-lck (by similarity).  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Name=1;  
 CC Event=Alternative splicing; Named Isoforms=2;  
 CC Name=2;  
 CC IsoId=P06332-1; Sequence=Displayed;  
 CC Name=2; Synonyms=Brain-specific;  
 CC IsoId=P06332-2; Sequence=VSP\_002489;  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: M36850; AAA39401.1; -;  
 CC EMBL: M13816; AAA37267.1; -;  
 CC EMBL: X04836; CAA28539.1; -;  
 CC EMBL: M36851; AAA39402.1; -;  
 CC EMBL: M17080; AAA37403.1; -;  
 CC EMBL: M17078; AAA37403.1; JOINED.  
 CC EMBL: AC002397; AAC36010.1; -;  
 CC EMBL: BC039137; AAH39137.1; -;  
 CC PIR: A02110; RWMST4.  
 CC HSSP: P01730; 1MBR.  
 CC MCD; MGI:88335; C44.  
 CC GO: GO:0042101; C:T-cell receptor complex; ISS.  
 CC GO: GO:0015026; F:coreceptor activity; ISS.  
 CC GO: GO:0042289; F:MHC class II protein binding; ISS.  
 CC GO: GO:0006955; P:immune response; ISS.  
 CC GO: GO:0045086; P:positive regulation of interleukin-2 biosyn. . .; ISS.  
 CC GO: GO:0030217; P:T-cell differentiation; ISS.  
 CC GO: GO:0045058; P:T-cell selection; ISS.  
 CC GO: GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; ISS.  
 CC InterPro: IPR000973; CD4 TCAG.  
 CC InterPro: IPR007110; Ig-Like.

DR InterPro: IPR003596; Ig\_V.  
 DR Pfam: PF00047; Ig\_2.  
 DR PRINTS; PRO0062; CD4TCANTIGEN.  
 DR SMART; SM00406; IGV; 1.  
 DR PROSITE; PS00835; IG\_LIKE; 1.  
 KW Immunoglobulin domain; Transmembrane; Glycoprotein; T-cell;  
 KW Immune response; Repeat; Signal; Lipoprotein; Palmitate;  
 KW Alternative splicing;  
 FT SIGNAL 1 26  
 FT CHAIN 27 457  
 FT DOMAIN 27 394  
 FT TRANSMEM 395 417  
 FT DOMAIN 418 457  
 FT DOMAIN 27 128  
 FT DOMAIN 129 207  
 FT DOMAIN 208 317  
 FT DOMAIN 318 374  
 FT CARBOHYD 187 187  
 FT CARBOHYD 298 298  
 FT CARBOHYD 323 323  
 FT CARBOHYD 392 392  
 FT DISULFID 42 112  
 FT DISULFID 159 188  
 FT DISULFID 328 370  
 FT LIPID 418 418  
 FT LIPID 421 421  
 FT VARSPLIC 1 240  
 FT  
 SQ SEQUENCE 457 AA; 51296 MW; 1BD4V527CB00F33 CRC64;  
 Query Match 29.1%; Score 993.5; DB 1; Length 457;  
 Best Local Similarity 50.9%; Pred. No. 3,9e-56;  
 Matches 217; Conservative 66; Mismatches 120; Indels 23; Gaps 8;  
 QY 1 MNRGVPFRH-LLVLTQLALLPAAQGNKVLAKGDPVELCTAAGKSKSIQPMKNSNOI 59  
 DB 1 MCRALSLRLLLLLLQLSQLAVTQKTLVLKGEBSALPESSQKTLTFWKFPSDOR 60  
 QY 60 KILGNQ-SFLTKG--PSKLNDRADSRSLMNGNFPLLIKULKIDSDSTYICEVDOKE 116  
 DB 61 KILGQHGKGVLLRGSSPSQF--DRFDSKKGAWKSGPPLINKLKMDSQTYICELENRKE 119  
 QY 117 EYQLVFGLTANSDFTLQSGSLTLTLES--PGSSSVVOCRSBPNIGQKTLVSQSE 175  
 DB 120 EYELNWFKXTFBPGSILQSGSLTLTLDNSKVSNPETCKKIKKGVSGSKLSNLR 179  
 QY 176 LQDSGTWCTVLQNGKVEFKIDIVLAFQKASSIVYKKEGVOVESFPPLAFTVEKLTGS 235  
 DB 180 VQSDSPWNTCTVLQDKNMFGMTLSVLFQSTAITAYKSEGSABEPSPLNFAEE--NGW 237  
 QY 236 GEIMWQAEPAASSKSKWITFDLKNKEYSVKRVTQDPKLGKQLPLHLTPALPOYAGSG 295  
 DB 238 GEIMWQAEKDSFPQWISPSIRKKEYSVQKSTKDLQKELPLPLTKIPOVSLQAGSG 297  
 QY 296 NLTLEAKTGKLDHENVVMRATQLOKNLCEVWGPSPKMLSLKLENKAKYSKKE 355  
 DB 298 NLTLTLD--KGTLDQENVLVKVAQLNLTLCFVWGPSPKRLTLKQENQARVSEEO 355  
 QY 356 KPYVVLNPEAGWQCLSDSGVLLSENIKVLPTWSTPVPCEAPBPKSCKDTHTCBELLG 415  
 DB 356 KVVQVAVPETGLMQCLSGDKVKMDSRIOVL-----SRGNQVTFVLACVLG 402  
 QY 416 GPSVFL 421  
 DB 403 GSGFPL 408  
 RESULT 19  
 GC3M MOUSE STANDARD; PRT; 398 AA.  
 ID GC3M MOUSE  
 AC P03987;  
 DT 23-OCT-1986 (Rel. 02, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)



```

DR PDB; 1F11; 06-FEB-01.
DR PDB; 1F58; 29-DEC-99.
DR PDB; 1KCS; 24-JUL-02.
DR PDB; 1KCR; 11-MAY-02.
DR PDB; 25C8; 09-JUL-99.
DR MGD; MGI:96446; Igh-4.
DR InterPro; IPR007110; Ig-1ike.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1; 2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing; Transmembrane; 3D-structure.
FT NON_TER 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 110 HINGE.
FT DOMAIN 111 217 CH2.
FT DOMAIN 218 324 CH3.
FT DISUFID 27 82
FT DISUFID 102 102 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISUFID 104 104 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISUFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISUFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISUFID 138 198
FT CARBOHYD 174 174 N-LINKED (GLCNAC. . .).
FT DISUFID 244 302
FT TRANSLEM 340 357
FT DOMAIN 358 393 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 393 AA; 43386 MW; 4CC88343B7A1CE27 CRC64;

Query Match 27.2%; Score 928.5; DB 1; Length 393;
Best Local Similarity 57.2%; Pred. No. 4,4e-52;
Matches 167; Conservative 51; Mismatches 47; Indels 27; Gaps 3;

QY 379 LLESNKIVLPFWSPVPCAP-EPKSGCDKH-----TCPELL 414
DB 56 VLDSDLTSSSVTVSSPPSEVTGCVNAPASSTVKDKIYPRDCCPCICTVPEV- 114
QY 415 GGSVFLFPKPKDMLISRTPEVTCVVDVSHEDPEVKENMYDGVENVAKTKPREQ 474
DB 115 --SSVFIFPKPKDVLITITLPKTCVVDISKDPEVQSWVDVEVHTAQDREBQ 172
QY 475 YNSTYRVSVLTVLHODMLNGKEYCKVSKALPAPEKTIISAKGQPREPVYTLPPSR 534
DB 173 FNSFTFSVSLPIMHODMLNGKEYCKVNSAAPAPIEKTIISKTKGPKAPQVYTLPPPK 232
QY 535 DELTKQVSLTCLVKGFPSPDIWEMENSGOPENNYKTPPVLDOSGPFYKLTVDKS 594
DB 233 EQWAKDQVSLTCLMTTFDFPEDITVEMQGNOPENNYKNTPIVNTNGSYFVSKLVQKS 292
QY 595 RMOGVNFSVCSVNEALHNHYTKSLSPGLQIDETCAEADGELDGLWTT 646
DB 293 NMEAGNFTCSVLEHGLHNHTEKSLSHSPGLQIDETCAEADGELDGLWTT 344

RESULT 21
GC RABIT STANDARD; PRT; 323 AA.
AC POL870;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma chain C region.
OS Oryctolagus cuniculus (Rabbit).
OC Oryctolagus cuniculus; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=8403030; PubMed=6313520;
RA Bernetein K.E., Alexander C.B., Mage R.G.;

```

```

RT "Nucleotide sequence of a rabbit IgG heavy chain from the recombinant
RT F-I haplotype.";
RN Immunogenetics 18:387-397(1983).
RL (2)
RP SEQUENCE OF 1-128.
RX MEDLINE=76135469; PubMed=1243651;
RA Pratt D.M., Mole L.E.;
RT "Sequence studies on the constant region of the Fd sections of rabbit
RT immunoglobulin G of different allotype.",
RN Biochem. J. 151:337-349(1975).
RL (3)
RP SEQUENCE OF 88-266 FROM N.A.
RX MEDLINE=83299917; PubMed=6193512;
RA Marcens C.L., Moore K.W., Steinmetz M., Hood L., Knight K.L.;
RT "Heavy chain genes of rabbit IgG: isolation of a cDNA encoding gamma
RT heavy chain and identification of two genomic C gamma genes.",
RN Proc. Natl. Acad. Sci. U.S.A. 79:6018-6022(1982).
RL (4)
RP SEQUENCE OF 132-161.
RX MEDLINE=70110015; PubMed=5461106;
RA Fruchter R.G., Jackson S.A., Mole L.E., Porter R.R.;
RT "Sequence studies of the Fd section of the heavy chain of rabbit
RT immunoglobulin G.",
RN Biochem. J. 116:249-259(1970).
RL (5)
RP SEQUENCE OF 129-131 AND 155-322.
RA Hill R.L., Lebovitz H.E., Fellows R.E. Jr., Delaney R.;
RT (in) Kildander J. (eds.);
RN Gamma globulins, Nobel symp. 3, pp.109-127, Almqvist and Wiksell,
RL Stockholm (1967).
CC -1- MISCELLANEOUS; Ref.1 sequence has the D12 allelic marker,
CC 104-Thr, and the E14 marker, 185-Thr. Ref.3 has the D11 and E15
CC markers and Ref.5 the E15 marker.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M16436; AAA31289.1; -
DR PIR; A91749; GHRB.
DR HSSP; P01857; IFC1.
DR InterPro; IPR007110; Ig-1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1; 2.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON_TER 1
FT DOMAIN 1 96
FT DOMAIN 114 213
FT DOMAIN 222 318
FT VARIANT 104 104
FT VARIANT 185 185
FT CONFLICT 48 48
FT CONFLICT 71 71
FT CONFLICT 144 144
FT CONFLICT 173 173
FT CONFLICT 187 187
FT CONFLICT 201 201
FT CONFLICT 218 218
FT CONFLICT 233 233
FT CONFLICT 246 246
FT CONFLICT 256 256
FT CONFLICT 260 260
FT CONFLICT 266 266
FT CONFLICT 280 280
FT IG-LIKE 1.
FT IG-LIKE 2.
FT T->M (IN D11 MARKER).
FT T->A (IN E15 MARKER).
FT N->E (IN REF. 2).
FT V->VPV (IN REF. 2).
FT Q->E (IN REF. 3 AND 4).
FT N->D (IN REF. 5).
FT Q->E (IN REF. 5).
FT E->G (IN REF. 5).
FT N->D (IN REF. 5).
FT N->D (IN REF. 5).
FT Y->W (IN REF. 5).

```

FT CONFLICT 284 284 N -> S (IN REF. 5).  
 SQ SEQUENCE 323 AA; 35404 MW; 69E8BA118D579A8B CRC64;  
 Query Match 26.7%; Score 910; DB 1; Length 323;  
 Best Local Similarity 65.7%; Pred. No. 5.1e-51;  
 Matches 174; Conservative 31; Mismatches 44; Indels 16; Gaps 5;

QY 375 SGQVLEENIKYLPWSTPVC---PAPB-----PKSCDKTHTC--PELLGGPPEVF 420  
 DB 60 SGVLSLSVSVTS-SQPTVCNVAHPATNTKDKTVAPTSCSR-TPEPELLGGPPEVF 117  
 QY 421 LPPPKKDTLMISRTPEVTCVVDVSHDEPEVKEMVYDGYEVHNAKTPREEQYNSTR 480  
 DB 118 IFPPKPKDTLMISRTPEVTCVVDVSGDPEVQFTWYINNOVTRAPRLREQQFNSTR 177  
 QY 481 VVSULTVHOMLNKEKCKVSNKALPAPTEKTSKAKGQPREPQVYTLPPSRDELTKN 540  
 DB 178 VVSTLPITHQDMLRKKEFKCKVHNKALPAPTEKTSKAKGQPREPQVYTLPPSRDELTKN 237  
 QY 541 QVSLTCLVKGFPSPDIAYEWESNGQPENNYKTPPVLDSDGSFPLYSKLTVDKSRMOGN 600  
 DB 238 VSLTCLMNGFPSPDISIVEMKNGKAEDNYKTPPAVLDSGSYFLYNKLSVPTSEMRGD 297  
 QY 601 VFSCSVMEALHNHYTKSLSPG 625  
 DB 298 VFTCSVMHEALHNHYTKSISRSPG 322

RESULT 22  
 GCAM\_MOUSE STANDARD; PRT; 399 AA.  
 ID GCAM\_MOUSE P01865;  
 AC P01865;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-AUG-1991 (Rel. 19, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig gamma-2A chain C region, membrane-bound form.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NX NCBI\_TaxId=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=82222190; PubMed=6283537;  
 RA Yamawaki-Kataoka Y., Nakai S., Miyata T., Honjo T.;  
 RT "Nucleotide sequences of gene segments encoding membrane domains of  
 RT immunoglobulin gamma chains";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:2623-2627(1982).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Membrane-bound;  
 CC IsoId=P01865-1; Sequence=Displayed;  
 CC Name=Secreted;  
 CC IsoId=P01864-1; Sequence=External;  
 CC Note=Probably the major isoform;  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; J00471; AAB59661.1; ALT\_INIT.  
 DR PIR; A02154; G2MSAM.  
 DR PDB; 1KB5; 06-APR-98.  
 DR PDB; 1YEE; 15-OCT-97.  
 DR MCD; MGI:96443; Igh-1.  
 DR InterPro; IPR007110; Igh-1-like.  
 DR InterPro; IPR003597; Igh\_c1.  
 DR InterPro; IPR003006; Igh\_MHC.  
 DR Pfam; PF00047; Igh; 2.

DR SMART; SM00407; IGC1; 2.  
 DR PROSITE; PS50835; IG\_LIKE; 3.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;  
 FT Transmembrane; Alternative splicing; 3D-structure; Repeat.  
 FT NON TER 1 1  
 FT DOMAIN 6 98  
 FT DOMAIN 121 220  
 FT DOMAIN 229 325  
 FT DISULFID 15 15  
 FT DISULFID 27 82  
 FT DISULFID 107 107  
 FT DISULFID 110 110  
 FT DISULFID 112 112  
 FT DISULFID 144 204  
 FT DISULFID 250 308  
 FT TRANSMEM 346 363  
 FT DOMAIN 364 399  
 FT CARBOHYD 180 180  
 SQ SEQUENCE 399 AA; 44020 MW; 4C38138BFAED3FF0 CRC64;

Query Match 26.4%; Score 902.5; DB 1; Length 399;  
 Best Local Similarity 49.0%; Pred. No. 2e-50;  
 Matches 180; Conservative 45; Mismatches 87; Indels 55; Gaps 5;

QY 285 PQLPQVAGSGNLTLEAKTKLHDEV-----NLVYMRATQLOKNTCEVWGPTSPKLM 339  
 DB 34 PEVTLTNSSGSLSSGVHTPPAVLQSDLYTLSSVTVTSSPQSITCVNAPAS---- 89  
 QY 340 LSLTKENKAKVSKREKPVVWVLPNPEAGMOCILSDSGVLLSNIKVLPTWSTPVCAP 399  
 DB 90 -----STKVKD-----KIEPRGPTIKPCP-- 108  
 QY 400 EPKSCDKTHTCPPELGGPSVFLPPPKKDTLMISRTPEVTCVVDVSHDEPEVKEMVYD 459  
 DB 109 -PCCK-----PAPNLGGPSVFIPEPKIKDYLMSISPIVTCVVDVSEDDPVQISWVN 163  
 QY 460 GVEVHNAKTPREEQYNSTRVSVVLTVLHODMNGEKYCKSNKALPAPTEKTSKAK 519  
 DB 164 NVEVHTAQDTQTHREDVNSTLRVVSALPIQHDWSSGKEPKCKVNNKQLPAPTEKTSKPK 223  
 QY 520 GQPREPQVYTLPPSRDELTKNOVSLTCLVKGFPSPDIAYEWESNGQPENNYKTPPVLD 579  
 DB 224 GSVAPAPVYVLPPEEEMTKQVTLCTGVTDPMEDLYVEMTNKGKELAYKTEPVLDS 283  
 QY 580 DGSFLLYSKLTVDKSRMOGNVPSGSVMHEALHNHYTKSLSPGLQDDETCAGADGE 639  
 DB 284 DGSYFMYSKLRVEKKNVERNSSCSYVHEGLNHHHTTKSFSTRPGLDLDVCAEADGE 343  
 QY 640 LDGLWTT 646  
 DB 344 LDGLWTT 350

RESULT 23  
 GC2\_CAVPO STANDARD; PRT; 329 AA.  
 ID GC2\_CAVPO P01862;  
 AC P01862;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE Ig gamma-2 chain C region.  
 OS Cavia porcellus (Guinea pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 NC Mammalia; Eutheria; Rodentia; Hystricognathi; Caviidae; Cavia.  
 NX NCBI\_TaxId=10141;  
 RN [1]  
 RP SEQUENCE OF 1-3.  
 RA Trischmann T.M.;  
 RL Submitted (APR-1975) to the PIR data bank.  
 RN [2]  
 RP SEQUENCE OF 4-68.  
 RX MEDLINE=71056471; PubMed=5538606;

Query Match	Best Local Similarity	Matches 164; Conservative	Score 879.5; DB 1; Length 329; Pred. No. 4,6e-49; Mismatches 39; Indels 7; Gaps 3
393 PWCAPAPKPSCKTKHTC--PELLGSGSVFLFPKPKTKLTMIISTPEVTCVVDVSHEDP 450	25.8%;	69.2%;	DB 1; Length 329; Pred. No. 4,6e-49; Mismatches 39; Indels 7; Gaps 3
96 PIRTPPBPTCPK--CPPENLGLGSPVFIIPPXPKDTLMTSLTPRYTCVVDVSDDEP 1522	25.8%;	69.2%;	DB 1; Length 329; Pred. No. 4,6e-49; Mismatches 39; Indels 7; Gaps 3
451 EVKENNVGVGVENHNAKTKPREEOVNSTYRVVSVYTLVADHOMLNGKCYKCKVSNKALPAP 510	25.8%;	69.2%;	DB 1; Length 329; Pred. No. 4,6e-49; Mismatches 39; Indels 7; Gaps 3
153 EVQFTVFDVKPVGNNETKPRVEQNTTTFRVESVLPIDHOMLNGKGEKCKVYNKALPAP 212	25.8%;	69.2%;	DB 1; Length 329; Pred. No. 4,6e-49; Mismatches 39; Indels 7; Gaps 3
511 IETIKISAKGQPREPOVYTLPPSRDELTKNQVSLTCLAKGSPEDIAVEMESGQPR--EN 568	25.8%;	69.2%;	DB 1; Length 329; Pred. No. 4,6e-49; Mismatches 39; Indels 7; Gaps 3
213 IETIKISKTGAPRMPVYTLPPSRDELSTKSVYTCGLINFPADIHVENANSRVPYSEK 272	25.8%;	69.2%;	DB 1; Length 329; Pred. No. 4,6e-49; Mismatches 39; Indels 7; Gaps 3
569 NYKTTTPVLDSDGSFFLYSKLITVDKSRMOQGVSCSWMEALHNHYTKSLSPG 625	25.8%;	69.2%;	DB 1; Length 329; Pred. No. 4,6e-49; Mismatches 39; Indels 7; Gaps 3

RESULT 24	GC	EMBL	GENE	MOUSE	STANDARD;	PROT;	405 AA.
AC	P01867;						
DT	21-JUL-1986 (Rel. 01, Created)						
DT	01-AUG-1991 (Rel. 19, Last sequence update)						
DT	15-MAR-2004 (Rel. 43, Last annotation update)						
DE	IG gamma-2B chain C region, membrane-bound form.						
OS	Mus musculus (Mouse).						
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;						
OC	Mammalia; Euteheria; Rodentia; Scturognathi; Muridae; Murinae; Mus.						
OX	NCBI_TaxID=10090;						
RN	[1]						
RP	SEQUENCE OF 335-405 FROM N.A.						
RX	MEDLINE=8222190; Pubmed=6283537;						
RA	Yamawaki-Kataoka Y., Nakai S., Miyata T., Honjo T.;						
RT	"Nucleotide sequences of gene segments encoding membrane domains of						
RT	immunoglobulin gamma chains."						
RL	Proc. Natl. Acad. Sci. U.S.A. 79:2623-2627(1982).						
RN	[2]						
RP	SEQUENCE OF 335-378 FROM N.A.						
RX	MEDLINE=82115295; Pubmed=6799207;						
RA	Rogers U., Choi E., Souza L., Carter C., Word C.J., Kuehl M.,						
RT	Eisenberg D., Wall R.;						
RT	"Gene segments encoding transmembrane carboxyl termini of						
RT	immunoglobulin gamma chains."						
RL	Cell 26:19-27(1981).						
CC	-1- ALTERNATIVE PRODUCTS:						
CC	Event=Alternative splicing; Named isoforms=2;						
CC	Name=Membrane-bound;						
CC	Isoid=P01867-1; Sequence=Displayed;						
CC	Name=Secreted;						
CC	Isoid=P01866-1; Sequence=External;						
CC	Note=May be the major isoform;						
CC	-1- PM: DISULFIDE BONDS BY HOMOLOG WITH OTHER IG GAMMA CHAINS.						
CC	-1- MISCELLANEOUS: The sequence of residues 1-335 is assumed to be						
CC	identical with the corresponding region of the secreted form.						
CC	-1- MISCELLANEOUS: The a allele sequence is shown.						
CC	-1- SIMILARITY: Contains 3 immunoglobulin-like domains.						
CC	-----						
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration						
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -						
CC	the European Bioinformatics Institute. There are no restrictions on its						
CC	use by non-profit institutions as long as its content is in no way						
CC	modified and this statement is not removed. Usage by and for commercial						
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>						
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).						
CC	-----						
CC	EMBL; J00462; AAB59659.1; ALT_INIT.						
DR	PIR; C02154; GZMSBM.						
DR	PDB; 1C1C; 11-MAR-03.						
DR	MGI; MGI:96445; Igh-3.						
DR	InterPro; IPR007110; Ig_1like.						
DR	InterPro; IPR003597; Ig_cl.						
DR	InterPro; IPR003006; Ig_MHC.						
DR	Pfam; PF00047; Ig; 3.						
DR	SMART; SM00407; IGH1; 2.						
DR	PROSITE; PS50835; IG_LIKE; 3.						
DR	PROSITE; PS00290; IG_MHC; 1.						
KW	Immunoglobulin domain; Immunoglobulin C region; Transmembrane;						
KW	Alternative splicing; 3D-structure; Repeat.						
FT	NON TER	1					
FT	DOMAIN	6	98				
FT	DOMAIN	127	226	IG-LIKE 1.			
FT	DOMAIN	235	331	IG-LIKE 2.			
FT	DISULFID	15	15	IG-LIKE 3.			

```

FT DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 118 118 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 150 210
FT DISULFID 256 314
FT TRANSMEM 352 369
FT DOMAIN 370 405 POTENTIAL.
SQ SEQUENCE 405 AA; 44330 MW; 89B3CF0A9B6D9FA CRC64;

Query Match 25.8%; Score 879.5; DB 1; Length 405;
Best Local Similarity 47.7%; Pred. No. 56-49; Indels 53; Gaps 6;
Matches 176; Conservative 54; Mismatches 86;

OY 285 PQALPQVAGSGLTLALBA---KTGKLHQLVNLVNRATQLQKLTCEVGPRTSLKM 339
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 34 PESTVTNNSSGLSSVHTFPALQSLYTMSSSVTSMPQGTVCVAHPAS----- 89
OY 340 LSLKLENKAKYKREKPPVWLNPRAGMOCILSSGCVLBSNLIKLPMTSTVPVCPAP 399
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 90 -----STVVDKLEP-----SGPI-----STINPCP-- 110
OY 400 EPKSCDKHTHC--PELLGSPVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPEVKEMWY 457
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 111 ---PCKCHCKCPAPVLEGSPVFLPPKIDVLMSTPKVTCVVVDVSEDDPDVQISWF 167
OY 458 VDGVEVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKITSK 517
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 168 VNNVEVHTAQTQTHEDYNSITRVSTLPQHODMGRKFKCKNNKNDLSPRIERTISK 227
OY 518 AKGQPREQVYTLTPSRDELTKNQVSLTCLYKGFPYSDIAVWESNGPENNYKTPPVYL 577
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 228 IKGLVRAQVYVILPPPAEQLSRKQVSLTCLVGFPPGDISVEMTNGHTEENYKDTAPVL 287
OY 578 DSDGSFELYSLKLTVDKSRMOCNVFSCVMHEALHNHTOKSLSLPGLQDCECAEQD 637
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 288 DSDGSFELYSLKLTVDKSRMOCNVFSCVMHEALHNHTOKSLSLPGLQDCECAEQD 637
OY 638 GELDGLMTT 646
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 348 GELDGLMTT 356

RESULT 25
GC3_MOUSE STANDARD; PRT: 329 AA.
AC P22436;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE IG gamma-3 chain C region, secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_Taxid=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=85027161; PubMed=6092053;
RA Weis J.A., Word C.J., Rimm D., Der-Balan G.P., Martinez H.M.,
RA Tucker P.W., Blatner F.R.;
RT "Structural analysis of the murine IgG3 constant region gene.";
RL EMBL J. 3:2041-2046(1984).

-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC
CC EMBL, J00451; -; NOT_ANNOTATED_CDS.
CC
CC DR PIR; B02156; G3MGC.
CC
CC DR HSSP; P01857; 1PCL.
CC
CC DR InterPro; IPR007110; Ig-1like.

```

```

DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; Igcl; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Transmembrane; Alternate splicing.
FT NON TER 1 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 113 HINGE.
FT DOMAIN 114 223 CH2.
FT DOMAIN 224 327 CH3.
SQ SEQUENCE 329 AA; 36228 MW; F45827174182BAD6 CRC64;

Query Match 24.6%; Score 840.5; DB 1; Length 329;
Best Local Similarity 64.5%; Pred. No. 1.46-46;
Matches 151; Conservative 37; Mismatches 39; Indels 7; Gaps 2;

OY 394 VPCPAPPKSCDKHTHC--ELLGSPVFLPPPKKDTLMSRTPEVTCVVVDVSHEDPE 451
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 100 IPKSTPPGS-----SCPPGNILEGSPVFLPPPKDLMSTPKVTCVVVDVSEDDPD 154
OY 452 VKFNMVYDGVVHNAKTKPREEQYNSTYRVVSVLTVLHQDWLNGKEYKCKVSNKALPAPI 511
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 155 VHVSMFVDNKKVHTAMQPREAQYNSTFRVVSALPIQHODMGRKFKCKNNKALPAPI 214
OY 512 EKTISKAKGQPREQVYTLTPSRDELTKNQVSLTCLYKGFPYSDIAVWESNGPENNYK 571
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 215 ERTISKPKGRQQTQVYVILPPPREQMSKKVSLTCLVTFPSEASISVEMENGELEDDYK 274
OY 572 TTPVLDSDGSFELYSLKLTVDKSRMOCNVFSCVMHEALHNHTOKSLSLSPG 625
|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|:::|
DB 275 TTPVLDSDGSFELYSLKLTVDKSRMOCNVFSCVMHEALHNHTOKSLSPG 625

RESULT 26
GC1_RAT STANDARD; PRT: 326 AA.
AC P20759;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE IG gamma-1 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89232738; PubMed=3149946;
RA Breggemann M.;
RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL Gene 74:473-482(1988).
DR PIR; PS0017; PS0017.
DR HSSP; P01842; 7PAB.
DR InterPro; IPR007110; Ig-1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; Igcl; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON TER 1 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 112 HINGE.
FT DOMAIN 113 219 CH2.
FT DOMAIN 220 326 CH3.
FT DISULFID 27 82
FT DISULFID 102 102
FT DISULFID 106 106
FT DISULFID 109 109
INTERCHAIN (WITH A HEAVY CHAIN).
INTERCHAIN (WITH A HEAVY CHAIN).
INTERCHAIN (WITH A HEAVY CHAIN).

```

```

FT DISULFID 111 111 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 140 200
FT DISULFID 246 304 N-LINKED (GLCNAC.. ) (POTENTIAL).
FT CARBOHYD 176 176
SQ SEQUENCE 326 AA; 35946 MW; 013BAB45EF49B9DA CRC64;

Query Match 24.3%; Score 829; DB 1; Length 326;
Best Local Similarity 57.3%; Pred. No. 7.4e-46;
Matches 153; Conservative 42; Mismatches 56; Indels 16; Gaps 4;

QY 375 SCQVLESNIKVLPTW-STVPCCPAPEPKSCDKT-HTCPPELLGG-----PS 418
DB 59 SGLYLTLSVTSVTSPSGTVCNVANHPASSSTKVDKVIPIRACGGGCKRCICTGSEVSS 118
QY 419 VFLEPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYDGVFNAAKTKEPEEQYNSI 478
DB 119 VFLEPPKPKDVLITLTPKTCVVVDVSDPEVHFSWFDVDEVTAAQTRPEEQFNST 178
QY 479 YRVSVLTVLHOMLNGKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDEL 538
DB 179 FRVSSELPILHQMILNGRTCKVTSAAFPSPLEKTIKPEGRTOVPHVITMSPTKEMT 238
QY 539 KNOVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLTVDKSRMQQ 598
DB 239 QNEVSLTCMVKGFPYPPDIYVEMQNGQPQENYKNTPTMTDGSYFLYSKLVNKKKMQQ 298
QY 599 GNVFSGSVMEALHNHYTQKSLSLSPG 625
DB 299 GNVFTCSVLHDEGLHNHTKEKLSHSPG 325

RESULT 27
GCB_RAT ID_GCB_RAT STANDARD; PRT; 333 AA.
AC P20761;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig gamma-2B chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=89232738; PubMed=3149946;
RA Brueggemann M.;
RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";
RL Gene 74:473-482(1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
DB HSP; P01842; 7FAB.
DR Interpro: IPR007110; Ig-1like.
DR Interpro: IPR003597; Ig_c1.
DR Interpro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_3.
DR SMART: SM00407; IgC1; 2.
DR PROSITE: PS50835; IG_LIKE; 3.
DR PROSITE: PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON TER 1 1
FT DOMAIN 1 96 IG-LIKE 1.
FT DOMAIN 124 223 IG-LIKE 2.
FT DOMAIN 232 328 IG-LIKE 3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 80 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 112 112 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 115 115 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 147 207 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 253 311
SQ SEQUENCE 333 AA; 36497 MW; 55FB8B64D48D460A6 CRC64;

```

```

Query Match 24.2%; Score 826.5; DB 1; Length 333;
Best Local Similarity 57.1%; Pred. No. 1.1e-45;
Matches 157; Conservative 35; Mismatches 58; Indels 25; Gaps 4;

QY 375 SCQVLESNIKVLPTW-STVPCCPAPEPKS-----CDKTHTC-- 410
DB 59 SGLYLTLSV-TSSITPSGTVCNVANHPASSSTKVDKVERNRNGIGCHKRCPTCTCKKCV 117
QY 411 PELLAGPSVFLFPPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNWYDGVFNAAKTKE 470
DB 118 PELLAGPSVFLFPPKPKDTLITISQNAKTCVVVDVSEEDPDVQFSWFNVAEVTATQTP 177
QY 471 REEOVNSTYRVSVLTVLHOMLNGKEYCKVSNKALPAPIEKTISKAKGQPREPOVYTL 530
DB 178 REEOVNSTYRVSVSLPIQIDQMMSGKEFKCVNNKALPSPIEKTISKPKGLVAKPQVYVW 237
QY 531 PPSRDELTKNQVSLTCLVKGFPYPSDIAVEMESNGQPENNYKTTTPVLDSDGSFFLYSKLT 590
DB 238 GPPTGLTETQVSLTCLTSGFLPNDIGVEMTNGHIEKKYKNTPEPVMDSGSEFMYSKLN 297
QY 591 YDKSRMQGNVFCSVMEALHNHYTQKSLSLSPG 625
DB 298 VERSRWDSRAPFCVSVEHGLHNHVEKSIKSRPG 332

RESULT 28
GCB_RAT ID_GCB_RAT STANDARD; PRT; 329 AA.
AC P20762;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig gamma-2C chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
[1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88166903; PubMed=3127222;
RA Brueggemann M.; Delmasstro-Galife P.; Waldmann H.; Calabi F.;
RT "Sequence of a rat immunoglobulin gamma 2c heavy chain constant
region cDNA: extensive homology to mouse gamma 3."
RL Eur. J. Immunol. 18:317-319(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.1eb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DB HSP; P01842; 7FAB.
DR Interpro: IPR007110; Ig-1like.
DR Interpro: IPR003597; Ig_c1.
DR Interpro: IPR003006; Ig_MHC.
DR Pfam: PF00047; Ig_2.
DR SMART: SM00407; IgC1; 2.
DR PROSITE: PS50835; IG_LIKE; 3.
DR PROSITE: PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 113 HINGE.
FT DOMAIN 114 222 CH2.
FT DOMAIN 223 329 CH3.
FT DISULFID 15 15 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 27 82 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 111 111 INTERCHAIN (WITH A HEAVY CHAIN).

```



```

FT DISULFID 113 113 INTERCHAIN (WITH A HEAVY CHAIN) .
FT DISULFID 143 203
FT DISULFID 249 307
SO SEQUENCE 329 AA; 36571 MW; 5FCD7B933850773 CRC64;

Query Match 24.1%; Score 822.5; DB 1; Length 329;
Best Local Similarity 57.0%; Pred. No. 2e-45;
Matches 154; Conservative 47; Mismatches 50; Indels 19; Gaps 4;

Oy 375 SGQVLESNIV-LPTWST-PVPC-----PA-----PEPKSCDTHTCPELLG 415
Db 59 SGLYTLSSSVTPSPSTSSQVTVCSVAHPATKSNLIKRIEPRPRPRPTDICSCDDMLG 118
Oy 416 GPSVFLPPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNWYDVEVNAKTRPREQY 475
Db 119 RPSVFIFPPKPKDITLMITLTKTCVVVDVSEEDPQPSMFVDNVRVFTAQDPHEQL 178
Oy 476 NSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIETKISKAKGQPREPVYTLPPSRD 535
Db 179 NGTFRVSVSLTHIQHODMWSGKEPKCKVNNKDLPSIEKTIKSPGKARTPOVYTIIPPRE 238
Oy 536 ELTRQVSLTCLVKGFPYSDIAVEENSGQPENNYKTTTPVLDSDGSEFLYSKLTVDKSR 595
Db 239 QMSKRVSLTCMVTSFYPASISVEWENGELEODYKNTLPYLDSDSEYFLYSKLSVDTDS 298
Oy 596 WQGNVFSCSYVMEALHNHYTKSLSPG 625
Db 299 WMRDITCSYVMEALHNHYTKSLSPG 328

RESULT 29
GCL_MOUSE
ID GCL_MOUSE STANDARD; PRT: 324 AA.
AC P01868;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE IG gamma-1 chain C region secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80045036; Pubmed=115593;
RA Honjo T., Obata M., Yamawaki-Kataoka Y., Kataoka T., Kawakami T.,
RA Takahashi N., Mano Y.,
RT "Cloning and complete nucleotide sequence of mouse immunoglobulin
RT gamma 1 chain gene.";
RL Cell 18:559-568(1979).
RN [2]
RP SEQUENCE OF 76-324 FROM N.A. (MYELOMA PROTEIN MOPC 31C).
RX MEDLINE=8020559; Pubmed=6769752;
RA Obata M., Yamawaki-Kataoka Y., Takahashi N., Kataoka T., Shimizu A.,
RA Mano Y., Seلمان J.G., Peterlin B.M., Leder P., Honjo T.;
RT "Immunoglobulin gamma 1 heavy chain gene: structural gene sequences
RT cloned in a bacterial plasmid.";
RL Gene 9:87-97(1980).
RN [3]
RP SEQUENCE OF 70-322 FROM N.A. (MYELOMA PROTEIN MOPC 21).
RX MEDLINE=80012837; Pubmed=113776;
RA Rogers J., Clarke P., Saleer W.,
RT "Sequence analysis of cloned cDNA encoding part of an immunoglobulin
RT heavy chain.";
RL Nucleic Acids Res. 6:3305-3321(1979).
RN [4]
RP SEQUENCE (MYELOMA PROTEIN MOPC 21).
RX MEDLINE=78242288; Pubmed=98524;
RA Adecugbo K.;
RT "Evolution of immunoglobulin subclasses. Primary structure of a
RT murine myeloma gamma1 chain.";
RL J. Biol. Chem. 253:6068-6075(1978).
RN [5]

```

```

RP DISULFIDE BONDS (MOPC 21).
RX MEDLINE=73008889; Pubmed=5073237;
RA Saeki J., Milstein C.;
RT "The disulphide bridges of a mouse immunoglobulin G1 protein.";
RL Biochem. J. 126:837-850(1972).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named Isoforms=2;
CC Name-Secreted;
CC IsoId=P01868-1; Sequence=Displayed;
CC Note=May be the major isoform;
CC Name=Membrane-Bound;
CC IsoId=P01869-1; Sequence=External;
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.ebi.ac.uk/announcements
or send an email to license@ebi.ac.uk).
-----
DR EMBL; V00793; CAA24172.1; -
DR EMBL; V00793; CAA24173.1; -
DR EMBL; V00793; CAA24174.1; -
DR EMBL; V00793; CAA24175.1; -
DR EMBL; V00795; CAA24176.1; -
DR PIR; A02159; GIMS.
DR GlycoSuiteDB; P01868; -
DR MGD; MGI:96446; Igh-4.
DR InterPro; IPR007110; Igh-1like.
DR InterPro; IPR003597; Igh_c1.
DR InterPro; IPR003006; Igh_MHC.
DR Pfam; PF00047; Igh; 2.
DR SMART; SM00407; Igh1; 2.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Alternative splicing.
FT NON_TER 1
FT DOMAIN 1 97 CH1.
FT DOMAIN 98 110 HINGE.
FT DOMAIN 111 217 CH2.
FT DOMAIN 218 324 CH3.
FT DISULFID 27 82
FT DISULFID 102 102 INTERCHAIN (WITH A LIGHT CHAIN).
FT DISULFID 104 104 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 109 109 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 138 198
FT CARBOHYD 174
FT FT
FT DISULFID 244 302
FT MOD_RES 324 324 REMOVED POST-TRANSLATIONALLY.
FT CONFLICT 276 276 N -> D (IN REF. 3).
FT CONFLICT 278 278 N -> D (IN REF. 3).
SO SEQUENCE 324 AA; 35704 MW; A538812F3D1F2C93 CRC64;

Query Match 23.9%; Score 814.5; DB 1; Length 324;
Best Local Similarity 53.9%; Pred. No. 6.2e-45;
Matches 146; Conservative 51; Mismatches 47; Indels 27; Gaps 3;

Oy 379 LLESINIVLPTWSTVPCAP-EPKSCDKTH-----TCPELL 414
Db 56 VLQSDLYTLSSSVTPSPSPSEETVTCNVHPASTKIVPRDGCRCICITVPEV- 114
Oy 415 GGSVFLPPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNWYDVEVNAKTRPREQY 474
Db 115 --SSVFIFPPKPKDITLITLTKTCVVVDVSEEDPQPSMFVDNVRVFTAQDPHEQL 172
Oy 475 NSTYRVSVLTVLHODMLNGEKYCKVSNKALPAPIETKISKAKGQPREPVYTLPPSRD 534
Db 173 FNGTFRVSVSLTHIQHODMLNGEKYCKVNNKDLPSIEKTIKSPGKARTPOVYTIIPPRE 232

```

OY 535 DELTKNOVSLTCLYKGFPSDIAVEMESNGCPENNYKTPPVLDSDGSFELYSLTJYDKS 594  
DB 233 EQAKDKVSLTCTITDFFEDITYEMQNGPAPENYKNTOPIMTNGSYFYVTSKLVNOKS 292  
OY 595 RMOQGVNFGSCSVMEALHNHYTOKSLSLSPG 625  
DB 293 NMEAGNTFTCSVLHNEGJHNHTKESLSHSPG 323

RESULT 30  
GCA\_MOUSE STANDARD; PRT; 330 AA.  
ID GCA\_MOUSE  
AC P01863;  
DT 21-JUL-1986 (Rel. 01, Created)  
DT 21-JUL-1986 (Rel. 01, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig gamma-2A chain C region, A allele.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=81076554; PubMed=6777755;  
RA Sikorav J.-L., Auffray C., Rougeon F.;  
RT "Structure of the constant and 3' untranslated regions of the murine  
RT Balb/c gamma 2a heavy chain messenger RNA.";  
RL Nucleic Acids Res. 8:3143-3155 (1980).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=81198976; PubMed=6262729;  
RA Yamawaki-Kataoka Y., Miyata T., Honjo T.;  
RT "The complete nucleotide sequence of mouse immunoglobulin gamma 2a gene  
RT and evolution of heavy chain genes: further evidence for intervening  
RT sequence-mediated domain transfer.";  
RL Nucleic Acids Res. 9:1365-1381 (1981).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=81223894; PubMed=6787604;  
RA Olio R., Auffray C., Morchamps C., Rougeon F.;  
RT "Comparison of mouse immunoglobulin gamma 2a and gamma 2b chain genes  
RT suggests that exons can be exchanged between genes in a multigenic  
RT family.";  
RL Proc. Natl. Acad. Sci. U.S.A. 78:2442-2446 (1981).  
RN [4]  
RP MYELOMA PROTEIN MOPC 173.  
RX MEDLINE=74175517; PubMed=4831970;  
RA Bourgois A., Fougereau M., Rocca-Serra J.;  
RT "Determination of the primary structure of a mouse IgG2a  
RT immunoglobulin amino-acid sequence of the Fc fragment. Implications  
RT for the evolution of immunoglobulin structure and function.";  
RL Eur. J. Biochem. 43:423-435 (1974).  
RN [5]  
RP DISULFIDE BONDS.  
RX MEDLINE=73056887; PubMed=4565406;  
RA de Preval C., Fougereau M.;  
RT "Determination of the primary structure of a mouse gamma G2a  
RT immunoglobulin. Identification of the disulfide bridges.";  
RL Eur. J. Biochem. 30:452-462 (1972).  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; V00798; CAA24178.1; -  
CC PIR; A02152; GMSA.  
CC PDB; 1E4W; 12-JUL-01.

DR PDB; 1E4X; 12-JUL-01.  
DR PDB; 1MUU; 06-MAY-99.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003557; Ig\_C1.  
DR InterPro; IPR003006; Ig\_MHC.  
DR Pfam; PF00047; Ig1; 2.  
DR SMART; SM00407; IgC1; 2.  
DR PROSITE; PS50835; IG\_LIKE; 3.  
DR PROSITE; PS00290; IG\_MHC; 1.  
KW Immunoglobulin domain; Immunoglobulin C region; 3D-structure; Repeat.  
FT NON\_TER 1 1  
FT DOMAIN 6 98  
FT DOMAIN 121 220  
FT DOMAIN 229 325  
FT DISULFID 15 15  
FT DISULFID 27 82  
FT DISULFID 107 107  
FT DISULFID 110 110  
FT DISULFID 112 112  
FT DISULFID 144 204  
FT DISULFID 250 308  
FT MOD\_RES 330 330  
SQ SEQUENCE 330 AA; 36389 MW; B84361C545A6864 CRC64;  
REMOVED POST-TRANSLATIONALLY.  
Query Match 23.5%; Score 801.5; DB 1; Length 330;  
Best Local Similarity 46.8%; Pred. No. 4,3e-44;  
Matches 162; Conservative 44; Mismatches 85; Indels 55; Gaps 5;  
OY 285 PQALPYAGSGNLTALAEKTKLHQEV-----NLVVMRATOLQKNLTCEWGPSPKLM 339  
DB 34 PEPVTLTNMNGSLSSGVHTFPAPVLQSDLYTLSSSVTSTSTPMSGITCVAHAPAS----- 89  
OY 340 LSLKLENKAKYKSKRPVWNLPEAGMNCCLSDGQVLBSNITVLPWSTFPVCPAP 339  
DB 90 -----STKYDK-----XIEPFGPIKPCP-- 108  
OY 400 EPKSCDKHTCPPELGGPSVFLPPPKKDTLMISRTPEYTCVVVDVSHEDPEVKFMVYD 459  
DB 109 -PKKC-----PAPNLGGPSVFLPPPKIKDVLMSLPITCVVDVSEDDPDVQISFVN 163  
OY 460 GVEVHNAKTRPREQNTYRVVSVLTVLAHOMLNGEKVKCVSKALPAPIKTSISKAK 519  
DB 164 NVEVHTAQQTREHDYNTLRVVSALPIQHDMWSKEKCKVKNNDLPAPIRITISKPK 223  
OY 520 GQPREQVYTLPPSRDELTRNQVSLTCLVKGFPSDIAVEMESNGCPENNYKTPPVLD 579  
DB 224 GSVRAPOVYVLPPEPEEMTKQVTLTCVTDPEPDIYVEWTNGKTELNKYTEPVLD 283  
OY 580 DGSFPLYSKLTVDKSRMOQGVNFGSCSVMEALHNHYTOKSLSLSPG 625  
DB 284 DGSYFVYSKLRVEKKNWERNISGCVHNEGJHNHTTYSFKTPG 329

RESULT 31  
GCA\_RAT STANDARD; PRT; 322 AA.  
ID GCA\_RAT  
AC P20760;  
DT 01-FEB-1991 (Rel. 17, Created)  
DT 01-FEB-1991 (Rel. 17, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Ig gamma-2A chain C region.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=89222738; PubMed=3149946;  
RA Bruggemann M.;  
RT "Evolution of the rat immunoglobulin gamma heavy-chain gene family.";  
RL Gene 74:473-482 (1988).  
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.  
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: M13804; AAA4376.1; ALT\_INNT.  
 DR PIR: PS0019; PS0019.  
 DR HSP: P01842; 7FAB.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003597; Ig cl.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR Pfam: PF00047; Ig; 2.  
 DR SMART: SM00407; Igcl; 2.  
 DR PROSITE: PS50835; IG-LIKE; 3.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 KM Immunoglobulin domain; Immunoglobulin C region; Repeat.  
 FT NON TER 1 1  
 FT DOMAIN 6 98 IG-LIKE 1.  
 FT DOMAIN 115 212 IG-LIKE 2.  
 FT DOMAIN 221 317 IG-LIKE 3.  
 FT DISULFID 27 82  
 FT DISULFID 102 102 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT DISULFID 105 105 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT DISULFID 107 107 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT DISULFID 136 196  
 FT DISULFID 242 300  
 FT CARBOHYD 172 172  
 SQ SEQUENCE 322 AA; 35186 MW; BBAJ36A9DE01EDB CRC64; N-LINKED (GLCNAC. . .) (POTENTIAL).

Query Match 23.2%; Score 792; DB 1; Length 322;  
 Best Local Similarity 56.4%; Pred. No. 1.7e-43;  
 Matches 150; Conservative 39; Mismatches 59; Indels 18; Gaps 4;

QY 375 SGQVLSNRIKY-LPTNST-PVPCPAPKSCDKTH-----TCPELLGGPSV 419  
 DB 59 SGLTTLTSVTVPSSTSSQAVTCNVAFHASTVKDKIVRECNPGCGTSEV---SSV 115  
 QY 420 FLPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKFNWYVDGVEHNATKPREBOYNSTY 479  
 DB 116 FIFPKTKDVLTLTLTPKTCVVVDISQNDPEVFPSMIDVEVHTAQTAAPEKQNSTL 175  
 QY 480 RVSVVLTVLHODMLNGEKYCKVKSNKALPAPIEKTISAKQKPPROYTLTPSRDELTK 539  
 DB 176 RSVSELPVHDMVNGKTFKCKVNSGAFPAPIEKSIKPESTPRGPQVYTWAPPEKEMTQ 235  
 QY 540 NQVSLTCLVKGFGYPSDIAVEMESNGOPENNYKTPPVLDSCGFYSKLVDSRNQOG 599  
 DB 236 SQVSIITCVKGFPPDITTEMKMGQPOENYKNTPTMDTGSYFLYSKLVNKKETWQOG 295  
 QY 600 NVEFSCVWHEALAHNYTQKSLSLSPG 625  
 DB 296 NTFCTSVLHBEGLHNHHTKSLSHSPG 321

RESULT 32  
 GCAB\_MOUSE  
 ID\_GCAB\_MOUSE STANDARD; PRT; 335 AA.  
 AC P01864;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE 15 gamma-2a chain C region secreted form (B allele).  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 CX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6;  
 RX MEDLINE=82037861; PubMed=6170065;

RA Schreier P.H., Bothwell A.L.M., Mueller-Hill B., Baltimore D.;  
 RT "Multiple differences between the nucleic acid sequences of the  
 RT Igg2a and Igg2b alleles of the mouse."  
 RL Proc. Natl. Acad. Sci. U.S.A. 78:4495-4499(1981).  
 RN [2]  
 RP SEQUENCE.  
 RX MEDLINE=82037777; PubMed=6794027;  
 RA Dognin M.J., Lauwereys M., Stroberg A.D.;  
 RT "Multiple amino acid substitutions between murine gamma 2a heavy  
 RT chain Fc regions of Iga1 and Igb1 alleotypic forms."  
 RL Proc. Natl. Acad. Sci. U.S.A. 78:4031-4035(1981).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Secreted;  
 CC IsoId=P01864-1; Sequence=displayed;  
 CC Note=Probably the major isoform;  
 CC Name=Membrane-bound;  
 CC IsoId=P01865-1; Sequence=External;  
 CC -1- MISCELLANEOUS: The sequence differs from that of the a allele,  
 CC from BALB/c mice, at 15% of the positions.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

DR EMBL: J00479; -; NOT\_ANNOTATED\_CDS.  
 DR PIR: A02153; GZMSAB.  
 DR PDB: 1BOG; 23-JAN-99.  
 DR PDB: 1HH6; 26-JAN-01.  
 DR PDB: 1HH9; 24-JUL-03.  
 DR PDB: 1HI6; 08-FEB-01.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003597; Ig cl.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR Pfam: PF00047; Ig; 2.  
 DR SMART: SM00407; Igcl; 2.  
 DR PROSITE: PS50835; IG-LIKE; 3.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 KM Immunoglobulin domain; Immunoglobulin C region; Alternative splicing;  
 FT 3D-structure; Repeat. 1  
 FT DOMAIN 6 98 IG-LIKE 1.  
 FT DOMAIN 126 225 IG-LIKE 2.  
 FT DOMAIN 234 330 IG-LIKE 3.  
 SQ SEQUENCE 335 AA; 36596 MW; FA3382792CBBI3C6 CRC64;

Query Match 23.1%; Score 788; DB 1; Length 335;  
 Best Local Similarity 45.7%; Pred. No. 3.1e-43;  
 Matches 158; Conservative 53; Mismatches 85; Indels 50; Gaps 5;

QY 285 PQALPYAGSGNTLALIA-----KTKSLHGVNLVVMKATLOLQKNTLCEVGFSPKLM 339  
 DB 34 PEPTLTLMNSGLSSGVHTFPALLQSLYTLSSSVTSTWTPQITTCVVAHPAS---- 89  
 QY 340 LSLKLENKAKVSKREKPVWVNLNPEAGMCCLSDSQVLLSEIKYKLPMTSTVPCCAP 399  
 DB 90 -----STVDVKTIIP-----RVPTQN-----PCP-- 109  
 QY 400 EPKSCDKTHTCPELLGGPSVFLFPPPKDPTLMISRTPEVTCVVVDVSHEDPEVKFNWYVD 459  
 DB 110 -PHQRYVPCAPADLLGSPSFIFFPKIKDVLMLISLSTFMVTCVVVDSEDDPDVQISMFVN 168  
 QY 460 GVEVHNATKPREBOYNSTYRVSVLTVLHODMLNGEKYCKVKSNKALPAPIEKTISKAK 519  
 DB 169 NVEVHTAQQTGHRREDVNSTLRVVSALPIQHDWMSGKFKCKVNNRALLPSIEKTISKPR 228  
 QY 520 GQPREPVYTLTPSRDELTKNQVSLTCLVKGFGYPSDIAVEMESNGOPENNYKTPPVLD 579

Db 229 GPRAPQVYVLPPEAEEMTKKESLTCMTGFLPAILADWMTSGTEQNYKTAIVLDS 288  
 QY 580 DGSFPLYSKLTVDKSRMOCNVFSCVMEALHNHYTKSLSPG 625  
 Db 289 DGSYFMYSKLRVQKSTWERSLFPACSVHEVHLNHLTTKTSRLG 334

RESULT 33  
 GCB\_MOUSE  
 ID GCB\_MOUSE STANDARD; PRT; 336 AA.  
 AC P01866;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE 1g gamma-2b chain C region secreted form.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCBITaxID=10090;  
 RN (1)  
 RN SEQUENCE FROM N.A. (ALLEN A).  
 RX MEDLINE=80120716; PubMed=6766534;  
 RA Yamawaki-Kataoka Y., Kataoka T., Takahashi N., Obata M., Honjo T.;  
 RT "Complete nucleotide sequence of immunoglobulin gamma2b chain gene  
 RT cloned from newborn mouse DNA."  
 RL Nature 283:786-789(1980).  
 RN (2)  
 RN SEQUENCE FROM N.A. (MPC 11).  
 RX MEDLINE=80081501; PubMed=117548;  
 RA Tucker P.W., Marcu K.B., Slightner J.L., Blattner F.R.;  
 RT "Structure of the constant and 3' untranslated regions of the murine  
 RT gamma 2b heavy chain messenger RNA."  
 RL Science 206:1299-1303(1979).  
 RN (3)  
 RN SEQUENCE FROM N.A.  
 RX MEDLINE=80081502; PubMed=117549;  
 RA Tucker P.W., Marcu K.B., Newell N., Richards J., Blattner F.R.;  
 RT "Sequence of the cloned gene for the constant region of murine gamma  
 RT 2b immunoglobulin heavy chain."  
 RL Science 206:1303-1306(1979).  
 RN (4)  
 RN SEQUENCE FROM N.A. (ALLEN B).  
 RX MEDLINE=82173203; PubMed=6803173;  
 RA O'Lo R., Rougeon F.;  
 RT "Mouse immunoglobulin allotypes: post-duplication divergence of gamma  
 RT 2a and gamma 2b chain genes."  
 RL Nature 296:761-763(1982).  
 RN (5)  
 RN CARBOHYDRATE-LINKAGE SITE THR-105.  
 RP MEDLINE=94216359; PubMed=7512967;  
 RA Kim H., Yamaguchi Y., Maeda K., Matsunaga C., Yamamoto K.,  
 RA Irimura T., Takahashi N., Kato K., Arata Y.;  
 RT "O-glycosylation in hinge region of mouse immunoglobulin G2b."  
 RL J. Biol. Chem. 269:12345-12350(1994).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Potential).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=Secreted;  
 CC IsoId=P01866-1; Sequence=Displayed;  
 CC Note=May be the major isoform;  
 CC Name=Membrane-bound;  
 CC IsoId=P01867-1; Sequence=External;  
 CC -1- PTM: O-LINKED GLYCIN CONSISTS OF GAL-GALNAC DISACCHARIDE WHICH IS  
 CC MODIFIED WITH 2 SIALIC ACID RESIDUES.  
 CC -1- PTM: DISULFIDE BONDS BY HOMOLOG WITH OTHER IG GAMMA CHAINS.  
 CC -1- MISCELLANEOUS: Contains a allele sequence is shown.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.  
 DR HSP; S25057; G2MS11.  
 DR HSP; P01842; 7FAB.  
 DR InterPro: IPR007110; IG-like.  
 DR InterPro: IPR003597; IG cl.  
 DR InterPro: IPR003006; IG\_MHC.

DR Pfam; PF00047; Ig; 3.  
 DR SMART; SM00407; IGc1; 2.  
 DR PROSITE; PS00835; IG-Like; 3.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;  
 KW Alternative splicing; Repeat.  
 FT NON TER 1  
 FT DOMAIN 6 98  
 FT DOMAIN 127 226  
 FT DOMAIN 235 331  
 FT DISULFID 15 15  
 FT DISULFID 27 82  
 FT DISULFID 109 109  
 FT DISULFID 112 112  
 FT DISULFID 115 115  
 FT DISULFID 118 118  
 FT DISULFID 150 210  
 FT DISULFID 256 314  
 FT CARBOHYD 105 105  
 FT MOD RES 336 336  
 FT VARIANT 163 163  
 FT VARIANT 194 194  
 FT VARIANT 300 300  
 FT VARIANT 301 301  
 FT CONFLICT 25 25  
 FT CONFLICT 36 36  
 FT CONFLICT 239 239  
 SQ SEQUENCE 336 AA; 36658 MW; 7D879662607C356E CRC64;  
 Query Match 22.9%; Score 783.5; DB 1; Length 336;  
 Best local Similarity 45.7%; Pred. No. 6; Ie-43;  
 Matches 159; Conservative 52; Mismatches 84; Indels 53; Gaps 6;  
 OY 285 PQLPOLYAGSGNLTALFA-----KTGKLDQEVNLTVMRATOLQKNLTCEWGPSPKLM 339  
 Db 34 PESTVTWNSGSLSSVHTFPALLOSLGYTWSSSVPSSTWPSQVTCVSHAPAS----- 89  
 OY 340 LSLKLENKAKYKSKRPVWVLPNPKGMWQCILSDGQVLLSINIKVLTWSPVPCAP 399  
 Db 90 -----STVDKLEP-----SGPI-----STINCP-- 110  
 OY 400 EPKSCDKHTC--PELLGSPSVFLPPPKPKDTLMSRTPEVTCVVDVSHEDPEVFNMY 457  
 Db 111 ----PCKECHKCAPNLEGSPSVIFPPNKKDVLMLSTKRTCVVDVSEDDDDVQISMF 167  
 OY 458 VQGEVHNKATYKREEQVNSTYRVVSVLTVLHODVNLNGEYKCKVSNKALPAIEKTIK 517  
 Db 168 VNNVEVHTAQOTHRDYNSTIRVSTLPIQHDWMSGKEPKCKVNNKDLPSPIERTISK 227  
 OY 518 AKGQREPGVYTLPRERDELTKNOVSLTCLVGVGYSDIAYVESNGQPENNYKTPPVY 577  
 Db 228 IKGLVAPQVYVLPPEAEQSLKRDVSLTCLVGFNPGDLSVETWSGHTENYKOTAPVL 287  
 OY 578 DSDGSEFPLYSKLTVDKSRMOCNVFSCVMEALHNHYTKSLSPG 625  
 Db 288 DSDGSEFPLYSKLNMKTSKWEKTDSPFCNVRHGLKNYLYLKTTSRLG 335

RESULT 34  
 EPC\_MOUSE  
 ID EPC\_MOUSE STANDARD; PRT; 421 AA.  
 AC P06336; P01856;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1996 (Rel. 33, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE 1g epsilon chain C region.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NC NCBITaxID=10090;  
 RN (1)  
 RN SEQUENCE FROM N.A.  
 RP MEDLINE=84236092; PubMed=6329728;

```

RA Ishida N., Ueda S., Hayashida H., Miyata T., Honjo T.;
RT "The nucleotide sequence of the mouse immunoglobulin epsilon gene:
RT comparison with the human epsilon gene sequence.";
RL EMBO J. 1.1117-1123(1982).
RN [2]
RP REVISIONS.
RA Honjo T.;
RL Submitted (APR-1986) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE OF 34-421 FROM N.A.
RX MEDLINE=83117774; PubMed=6618553;
RA Liu F.-T., Albrandt K., Sutcliffe J.G., Katz D.H.;
RT "Cloning and nucleotide sequence of mouse immunoglobulin epsilon
RT chain cDNA.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:7852-7856(1982).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X01857; CAA25977.1; -.
DR EMBL; X01857; CAA25978.1; -.
DR PIR; A02144; EHMS.
DR PIR; A02145; EHMS.
DR HSSP; P01854; 1IGB.
DR InterPro; IPR007110; Ig-1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_4.
DR SMART; SM00407; Ig_c1.2.
DR PROSITE; PS50835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
DR KX Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON TER 1 1
FT DOMAIN 1 90 CH1.
FT DOMAIN 91 197 CH2.
FT DOMAIN 198 304 CH3.
FT DOMAIN 305 421 CH4.
FT DISULFID 23 75 BY SIMILARITY.
FT DISULFID 121 180 BY SIMILARITY.
FT DISULFID 226 285 BY SIMILARITY.
FT DISULFID 330 392 BY SIMILARITY.
FT CARBOHYD 43 43 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 72 72 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 84 84 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 95 95 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 238 238 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 365 365 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 421 AA; 47320 MW; 8F909E1F30A06B47 CRC64;
Query Match 10.8%; Score 368.5; DB 1; Length 421;
Best Local Similarity 27.3%; Pred. No. 2.2e-16;
Matches 121; Conservative 72; Mismatches 152; Indels 99; Gaps 21;

```

```

QY 380 LESNIKVLPTWSTVPCEAPPEKSCDKTHNCPBLLGSPVFLPFPKPKDTLMISRTPEVT 439
DB LAHTRR-----CPDHEPR-----GVITLILPSPLD-LYQNGAPRLT 225
QY 440 CVVVDV-SHEDPEAKFN-----WYDGVVHNNAKTKPREBOYNSTYRVSVLTV 487
DB CLVVDLSEKQVNTWNOEKTSVSASQW---TKHN-----NATTSITSLPV 272
QY 488 LHODMLNGEKYKCVSNALAPAEKTSKAKGPREPOVYTLPPSRDELTKNOVSLTCL 547
DB 273 VAKDWIBSGYQCLVDHPDPKPIVRSITTKPGRSAPEVVFPPEEE-SEDKKTLTCL 331
QY 548 VKGFYPSDIAVEMESNGQPENN--YKTPPYLDSGDS---FELYSKLTVDKSRMQQNVF 602
DB 332 IQNFPEDISIQWLDGGLTNSQSHSTTP-LKSGNSNGGFFIRLEVAKTMTQKQF 390
QY 603 SCVWHEALHN-HYTKSLSLSPG 625
DB 391 TCQVYHEALQPRKLEKTIKSTSLG 414
RESULT 35
EPC_RAT STANDARD; PRT; 429 AA.
AC P01855;
DT 21-JUN-1986 (Rel. 01, Created)
DT 21-JUN-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DB Ig epsilon chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A. (IMMUNOCYTOMA IR2).
RC STRAIN=LOU/C/MSL;
RX MEDLINE=83064537; PubMed=6292865;
RA Hellman L., Petersen U., Engstrem A., Karlsson T., Bennich H.;
RT "Structure and evolution of the heavy chain from rat immunoglobulin
RT E.";
RL Nucleic Acids Res. 10:6041-6049(1982).
RN [2]
RP SEQUENCE OF 168-342 FROM N.A. (MYELOMA IR162).
RX MEDLINE=83182019; PubMed=6820340.
RA Kindsvogel W.R., Reddy E.P., Moore J.M., Faust C.H. Jr.;
RT "A cloned cDNA probe for rat immunoglobulin epsilon heavy chain:
RT construction, identification, and DNA sequence.";
RL DNA 1:335-343(1982).
RN [3]
RP SEQUENCE OF 205-306 FROM N.A.
RX MEDLINE=82174576; PubMed=6803238;
RA Hellman L., Petersen U., Bennich H.;
RT "Characterization and molecular cloning of the mRNA for the heavy
RT (epsilon) chain of rat immunoglobulin E.";
RL Proc. Natl. Acad. Sci. U.S.A. 79:1264-1268(1982).
CC -1- SIMILARITY: Contains 4 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00744; AAA41379.1; ALT_INIT.
DR PIR; A93442; EHRT.
DR HSSP; P01854; 1IGE.
DR InterPro; IPR007110; Ig-1like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig_4.
DR SMART; SM00407; Ig_c1. 1.

```



```

QY 518 AKGPRE-----POVYTLPSRDEL-T-KNOVSLTCLVKGFPYSPDIAYEWESNGQ--PENN 569
DQ 323 ----PNEVHKHPAVYLLPAPAREQLNRESATVTCVLVGFSPADISVQWLGRQGLLPQEK 378
DQ 570 YKTPRPVLD--SDGSFFLYSKLTVDKSMQGNVPSGCVMEHAIHNHTQKSLSPQLQ 627
DQ 379 YVTSAPMPEPAPGPGYFTHSILTVTEEMNSGETYTCVAGHEALPHLVTERTVDS---- 434
QY 628 LDETCABQDGLDGLWTT 646
DQ 435 -TEGEVNAEBEGFENLWTT 452

RESULT 37
MUC_HUMAN STANDARD; PRT: 454 AA.
ID AC 21-JUN-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig mu chain C region.
GN IGHM.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxId=9606;
RN (1)
RP MEDLINE=90332450; PubMed=2115996;
RA Friedlander R.M., Nussenzweig M.C., Leder P.;
RT "Complete nucleotide sequence of the membrane form of the human IgM
RT heavy chain.", Res. 18:4278-4278 (1990).
RL Nucleic Acids Res. 18:4278-4278 (1990).
RN (2)
RP MEDLINE=75059123; PubMed=4803843;
RA Watanabe S., Barnikol H.U., Horn J., Berttram J., Hilschmann N.;
RT "The primary structure of a monoclonal IgM-immunoglobulin
RT type", subgroup H III. Architecture of the complete IgM-molecule.",
RL Hoppe-Seyler's Z. Physiol. Chem. 354:1505-1509 (1973).
RN (3)
RP MEDLINE=81066716; PubMed=6777162;
RA Milnesco E., Barnikol-Watanabe S., Barnikol H.U., Milnesco C.,
RA Hilschmann N.;
RT "The primary structure of the constant part of mu-chain-disease
RT protein BOT.", Eur. J. Biochem. 111:275-286 (1980).
RN (4)
RP MEDLINE=74005511; PubMed=4742735;
RA Putnam F.W., Florent G., Paul C., Shintzu A.;
RT "Complete amino acid sequence of the mu heavy chain of a human IgM
RT immunoglobulin.", Science 182:287-291 (1973).
RN (5)
RP MEDLINE=82059479; PubMed=6795593;
RA Rabbits T.H., Forrester A., Milstein C.P.;
RT "Human immunoglobulin heavy chain genes: evolutionary comparisons of
RT C mu, C delta and C gamma genes and associated switch sequences.",
RL Nucleic Acids Res. 9:4509-4524 (1981).
RN (6)
RP MEDLINE=299-387 AND 438-454 FROM N.A.
RA MEDLINE=81073306; PubMed=6777778;
RA Dolby T.W., Devuono J., Croce C.W.;
RT "Cloning and partial nucleotide sequence of human immunoglobulin mu
RT chain cDNA from B cells and mouse-human hybridomas.",
CC Proc. Natl. Acad. Sci. U.S.A. 77:6027-6031 (1980).
CC -1- MISCELLANEOUS: All 4 combinations of the S/G and V/G polymorphisms
CC at positions 192 and 216 have been observed in human mu chains.

```

```

CC -----
CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb.ch).
CC -----
DQ EMBL; X17115; CAA34971.1; ALT SEQ.
DQ EMBL; X57086; ? NOT_ANNOTATED_CDS.
DQ HSSP; P01857; IFC1.
DQ Genew; HGNC:5541; IGHM.
DQ MIM; 147020; -.
DQ DR GlycoSiteDB; P01871; -.
DQ DR GO; GO:0005624; C:membrane fraction; NAS.
DQ DR GO; GO:0003823; F:antigen binding; NAS.
DQ DR GO; GO:0006955; P:immune response; NAS.
DQ DR InterPro; IPR007110; Ig-like.
DQ DR InterPro; IPR003597; Ig cl.
DQ DR InterPro; IPR003006; Ig_MHC.
DQ DR Pfam; PF00047; Ig; 4.
DQ DR SMART; SM00407; IGH1; 3.
DQ DR PROSITE; PS50835; IG LIKE; 4.
DQ DR PROSITE; PS00290; IG MHC; 3.
DQ KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
DQ polymorphism.
FT FT NON_TER 1 1
FT FT DOMAIN 1 105 CH1.
FT FT DOMAIN 106 218 CH2.
FT FT DOMAIN 219 324 CH3.
FT FT DOMAIN 325 454 CH4.
FT FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN).
FT FT DISULFID 28 88 INTERCHAIN (WITH HEAVY CHAIN).
FT FT DISULFID 135 198 INTERCHAIN (WITH HEAVY CHAIN).
FT FT DISULFID 215 215 INTERCHAIN (WITH HEAVY CHAIN IN ANOTHER
FT FT DISULFID 245 304 SUBUNIT).
FT FT DISULFID 292 292 INTERCHAIN (WITH HEAVY CHAIN).
FT FT DISULFID 352 414 INTERCHAIN (WITH HEAVY CHAIN).
FT FT DISULFID 453 453 N-LINKED (GLCNAC. . .).
FT FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).
FT FT CARBOHYD 210 210 N-LINKED (GLCNAC. . .).
FT FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .).
FT FT CARBOHYD 280 280 N-LINKED (GLCNAC. . .).
FT FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .).
FT FT VARIANT 192 192 /FTid=VAR_000219.
FT FT VARIANT 192 192 S->G.
FT FT VARIANT 216 216 /FTid=VAR_003903.
FT FT VARIANT 216 216 V->G (in dbSNP:12365).
SQ SEQUENCE 454 AA; 49556 MW; 21EC72EAD56922E CRC64;

Query Match 10.5%; Score 358.5; DB 1; Length 454;
Best Local Similarity 27.6%; Pred. No. 1.1e-15;
Matches 83; Conservative 62; Mismatches 111; Indels 45; Gaps 10;

QY 332 GPTSPKLSLKLKLNKAKVSKREKPVVNLNPEAGMQLSDSGVLE-SNIKVLPW 390
DQ 173 GPTTYKVTSLTITLIESD-----WL---SQSMFTCAVDHRLFGQVASMCPDQ 219
QY 391 STVPYCPAPPEKSCDKTHTCPDELIGSPVLPFPKPKDTLMISTPEVTGVVDVSHDP 450
DQ 220 DTAL-----RVFAIPPS-PASIFLTSTKLTCLVTLDTTYD- 254
QY 451 EVKKNWYVDGVENNAKTKPREQVNSTYRVSVSLTVLHODMNLGKXKCVSKALPAP 510
DQ 255 SVTISWTRNGEAVKTHNINSESHPNATPSAVGASLCEDDMNSGEFFTCVHTTDLPS 314
QY 511 IETKISRAGQP-REPQVYTLPSRDEL-T-KNOVSLTCLVKGFPYSPDIAYEWESNGQP-- 566
DQ 315 LKQTSRPKGVAAHRPVYLLPAPAREQLNRESATITCVLGFSFPAVVFVQMGQGPLS 374

```

0Y	567	ENNKTPRPVDD--SDGSFFLIXSTYDKSKMOGNNVSCSWMEALHNHTOKSLSP	624
Db	375	PEKVTSAPMPEQAPGRYFAHSILTYSEEMWNTGETYCVAAHEALPNRYTERTVDKST	434
0Y	625	G 625	
Db	435	G 435	
RESULT 38			
EPC_HUMAN			
ID	EPC_HUMAN	STANDARD;	PRT; 428 AA.
AC	P01854;		
DT	21-JUL-1986	(Rel. 01, Created)	
DT	21-JUL-1986	(Rel. 01, Last sequence update)	
DT	10-OCT-2003	(Rel. 42, Last annotation update)	
DE	Ig epsilon chain C region.		
DE	Ig epsilon chain C region.		
GN	IGH.		
OS	Homo sapiens (Human).		
CC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
CC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.		
OX	NCBI_Taxid=9606;		
PN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=83168897; PubMed=6300763;		
RA	Seno M., Kurckawa T., Ono Y., Onda H., Sasada R., Igarashi K.,		
RA	Kikuchi M., Sugino Y., Nishida Y., Honjo T.;		
RT	"Molecular cloning and nucleotide sequencing of human immunoglobulin		
RT	epsilon chain cDNA.";		
RL	Nucleic Acids Res. 11:719-726(1983).		
RN	[2]		
RP	SEQUENCE FROM N.A., AND VARIANT LEU-359.		
RX	MEDLINE=83001945; PubMed=6288268;		
RA	Max E.E., Batcey J., Ney R., Kirsch I.R., Leder P.;		
RT	"Duplication and deletion in the human immunoglobulin epsilon genes.,"		
RL	Cell 29:691-699(1982).		
RN	[3]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=84236029; PubMed=6234164;		
RA	Flanagan J.G., Rabbits T.H.;		
RT	"The sequence of a human immunoglobulin epsilon heavy chain constant		
RT	region gene, and evidence for three non-allelic genes.";		
RL	EMBO J. 1:655-660(1982).		
RN	[4]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=84207910; PubMed=6327276;		
RA	Ueda S., Nakai S., Nishida Y., Hisejima H., Honjo T.;		
RT	"Long terminal repeat-like elements flank a human immunoglobulin		
RT	epsilon pseudogene that lacks introns.";		
RL	EMBO J. 1:1539-1544(1982).		
RN	[5]		
RP	PRELIMINARY SEQUENCE (MYELOMA PROTEIN ND).		
RA	Bernich H.H., Johansson S.G.O., von Bahr-Lindstrom H.;		
RA	(in) Bach M.K. (eds.).;		
RT	Immediate hypersensitivity: modern concepts and developments, pp.1-36,		
RL	Marcel Dekker, New York (1978).		
RN	[6]		
RP	SEQUENCE OF 1-40; 68-114 AND 427-428 FROM N.A.		
RX	MEDLINE=83065234; PubMed=6815556;		
RA	Kerten J.H., Moigaard H.V., Houghton M., Derbyshire R.B., Viney J.,		
RA	Bell L.O., Gould H.J.;		
RT	"Cloning and sequence determination of the gene for the human		
RT	immunoglobulin epsilon chain expressed in a myeloma cell line.";		
RL	Proc. Natl. Acad. Sci. U.S.A. 79:6661-6665(1982).		
RN	[7]		
RP	3D-STRUCTURE MODELING.		
RX	MEDLINE=87089848; PubMed=3796618;		
RA	Padlan E.A., Davies D.R.;		
RT	"A model of the FC of immunoglobulin E.";		
RL	Mol. Immunol. 23:1063-1075(1986).		
CC	-1- SIMILARITY: Contains 4 immunoglobulin-like domains.		
CC	-----		
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration		

CC	between the Swiss Institute of Bioinformatics and the EMBL outstation
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.1ab-sib.ch/announce/">http://www.1ab-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@sib-sib.ch">license@sib-sib.ch</a> ).
CC	-----
DR	EMBL; L00022; AAB59424.1; ALT_INIT.
DR	PIR; A22771; EHHU.
DR	PDB; 1IGE; 15-JUL-92.
DR	PDB; 1EP5; 30-JAN-02.
DR	PDB; 1G84; 16-MAY-01.
DR	PDB; 1OOV; 18-SEP-02.
DR	Genew; HGNC:5522; IGHE.
DR	MIM; 147180; -.
DR	GO; GO:0003823; P:antigen binding; NAS.
DR	GO; GO:0006955; P:immune response; NAS.
DR	InterPro; IPR007110; Ig-like.
DR	InterPro; IPR003597; Ig-cl.
DR	InterPro; IPR003006; Ig_MHC.
DR	Pfam; PF00047; Ig; 4.
DR	SMART; SM00407; IGc1; 4.
DR	PROSITE; PS00835; IG_LIKE; 4.
DR	PROSITE; PS00290; IG_MHC; 3.
KW	Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW	3D-structure; Repeat.
FT	NON TER 1
FT	DOMAIN 6
FT	DOMAIN 103
FT	DOMAIN 112
FT	DOMAIN 210
FT	DOMAIN 214
FT	DOMAIN 318
FT	DOMAIN 324
FT	DOMAIN 423
FT	DISULFID 14
FT	DISULFID 14
FT	DISULFID 15
FT	DISULFID 105
FT	DISULFID 29
FT	DISULFID 85
FT	DISULFID 121
FT	DISULFID 121
FT	DISULFID 135
FT	DISULFID 193
FT	DISULFID 209
FT	DISULFID 239
FT	DISULFID 299
FT	DISULFID 345
FT	DISULFID 405
FT	CARBOHYD 21
FT	CARBOHYD 21
FT	CARBOHYD 49
FT	CARBOHYD 49
FT	CARBOHYD 99
FT	CARBOHYD 99
FT	CARBOHYD 146
FT	CARBOHYD 146
FT	CARBOHYD 252
FT	CARBOHYD 252
FT	CARBOHYD 275
FT	CARBOHYD 275
FT	VARIANT 359
FT	VARIANT 359
FT	STRAND 110
FT	STRAND 110
FT	STRAND 113
FT	STRAND 117
FT	HELIX 122
FT	HELIX 124
FT	TURN 125
FT	TURN 125
FT	STRAND 130
FT	STRAND 141
FT	STRAND 146
FT	STRAND 150
FT	STRAND 155
FT	STRAND 155
FT	STRAND 159
FT	STRAND 161
FT	STRAND 165
FT	STRAND 166
FT	TURN 168
FT	TURN 169
FT	STRAND 172
FT	STRAND 181
FT	HELIX 182
FT	HELIX 186
FT	TURN 187
FT	TURN 188
FT	STRAND 192
FT	STRAND 196
FT	TURN 198
FT	TURN 199
FT	TURN 201
FT	TURN 202
FT	STRAND 205
FT	STRAND 207
FT	STRAND 218
FT	STRAND 222
FT	TURN 226
FT	TURN 229
FT	TURN 231
FT	TURN 232
FT	STRAND 236
FT	STRAND 242
FT	STRAND 252
FT	STRAND 252
FT	STRAND 255
FT	STRAND 257
FT	TURN 258
FT	TURN 259
FT	STRAND 260
FT	TURN 260
FT	TURN 265
FT	TURN 266



```
FT STRAND 271 274
FT TURN 275 276
FT STRAND 277 285
FT HELIX 288 293
FT TURN 294 294
FT STRAND 297 302
FT TURN 304 305
FT STRAND 310 314
FT STRAND 322 322
FT STRAND 325 329
FT HELIX 334 337
FT STRAND 340 348
FT STRAND 350 351
FT STRAND 356 360
FT TURN 366 369
FT STRAND 374 374
FT STRAND 377 378
FT TURN 380 381
FT STRAND 384 385
FT STRAND 387 393
FT HELIX 394 398
FT TURN 399 400
FT STRAND 404 408
FT STRAND 418 421
SQ SEQUENCE 428 AA; 47019 MW; 25C4CA072AAS58A0 CRC64;

Query March 10.4%; Score 356; DB 1; Length 428;
Best Local Similarity 26.6%; Pred. No. 1.4e-15;
Matches 118; Conservative 69; Mismatches 154; Indels 102; Gaps 18;

QY 267 TQDPKLT-----OMGKLLPLHLT-----LPOALPOVAGSGL---TLALEKTKGL 308
DB 3 TQSPSVFLPTCCCKNIPSNATSVTLGLATGYFEPVAVVTMDTGLNTMTLPTATITLT 62
QY 309 ---HOENVLVVRATOLQKNLTCEV-WGPTSPKMLSLKLENKAKVSKRE---KPVVVL 361
DB 63 SGHVAITSLTLVSGAMAKOMFTCRVAHPPSSTDW-----VDNKTFVSCSDPFTPTVKIL 117
QY 362 N-----PEAGMWOCLLS-----DSGOVLSSNKKVLPWTGTPPCAPAE 400
DB 118 QSSCDGGGHPPTTQLCLTVSGYTPGTINITWLEDDGYMDV---LSTASTOGBELAS 173
QY 401 PKS-----CDKHTTC-----DELGGPSPVFLFPPPKDITL 430
DB 174 TQSELTLSQKMLSDRTTCQVYOGHTFEDSTKKCADSNR---GVSAYLSRSPFD-L 229
QY 431 MISRTPEVTVVVDVSHEDPEVKENYVGVVHNAKTKPREEOYNSTYRVSVYLVLAHQ 490
DB 230 FIRKSPITCLVVDLAPSKGTVNLTWBRASGKPVVHSTRKEKQNGTLVTSLLPVQTR 289
QY 491 DMLNGKEYCKVSKNKALEPAPIEKITSKAKGQPREPOVYTL-----PPSRDELTKNOVSLT 545
DB 290 DWIGETVQCRVTHHLPRALMRSTTKTSGRPAPEVVAFAFATPEWPGSRDRT-----LA 344
QY 546 CLVNGEYPSDIAVEMESN---GQPENNYKTPPVLDSDSPFLYSKLTJYDKSRMOQANVES 603
DB 345 CLIQNFMEDISVQMLHNEVOLPDARHSTTTPRKTKSGGFVFSLEVTYRAEMQKDEFI 404
QY 604 CSVNHGALAHNYT-QKSLSLSPG 625
DB 405 CRAVHEAASPQTYQRAVSVNPG 427

RESULT 39
MUCM RABIT STANDARD; PRT; 479 AA.
AC P04221;
DT 20-MAR-1987 (Rel. 04, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 19 mu chain C region membrane-bound form.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
RX NCBI_Taxid=9986;
RN [1]
RP SEQUENCE FROM N.A. (A2 ALLOTYPED).
RX MEDLINE=84088930; PubMed=6418803;
RA Bernstein K.E., Alexander C.B., Reddy E.P., Mage R.G.;
RT "Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain
RL J. Immunol. 132:490-495(1984).
CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Comment-During differentiation, B lymphocytes switch from
CC expression of isoform Membrane-bound to isoform Secreted;
CC Name=Membrane-bound;
CC IsoId=P04221-1; Sequence=Displayed;
CC Name=Secreted;
CC IsoId=P03988-1; Sequence=External;
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
```

EMBL; K01357; AAA31293.1; -  
PIR; A02165; MHRBM.  
DR HSSP; P01842; 7PAB.  
DR InterPro; IPR007110; Ig-like.  
DR InterPro; IPR003597; Ig cl.  
DR InterPro; IPR003006; Ig\_MHC.  
DR Pfam; PF00047; Ig; 4.  
DR SMART; SM00407; Igcl; 2.  
DR PROSITE; PS50835; IG\_LIKE; 4.  
DR PROSITE; PS00290; IG\_MHC; 3.  
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;  
KW Alternative splicing; Transmembrane.  
FT NON\_TER 1 1  
FT DOMAIN 1 106 CH1.  
FT DOMAIN 107 222 CH2.  
FT DOMAIN 223 327 CH3.  
FT DOMAIN 328 458 CH4.  
FT TRANSMEM 459 476  
FT DISULFID 14 14  
FT DISULFID 28 90  
FT DISULFID 137 200  
FT DISULFID 219 219  
FT DISULFID 249 308  
FT DISULFID 296 296  
FT DISULFID 356 418  
FT CARBOHYD 46 46  
FT CARBOHYD 114 114  
FT CARBOHYD 212 212  
FT CARBOHYD 261 261  
FT CARBOHYD 277 277  
FT CARBOHYD 284 284  
SQ SEQUENCE 479 AA; 52351 MW; 689C637A47BE19FC CRC64;

Query March 10.4%; Score 355.5; DB 1; Length 479;  
Best Local Similarity 24.8%; Pred. No. 1.6e-15;  
Matches 122; Conservative 81; Mismatches 176; Indels 113; Gaps 21;

QY 223 FPL-----AFTVEKLTSGGELMWAQERASSKSNITFDLKNKEVSVKRVYVTDQPKLQMGKK 277  
DB 9 YPLVSCGALTGDGLVAMGCLARDPLSSVTFSW-SFK-NSSEISSRTVTFPPVVKRGDK 66  
QY 278 -----LPLHLTPQAL-----PQYA 292  
DB 67 YMATSQVLVPSKDVLTQSTEEYLVCKYQHSNSNDRVSPVDSLPENVSFVTFPPRDSFS 126  
QY 293 GSGNLTALAEAK-TGKLHGEVNLVVMRATQ-----LQKLTCEVWGPPTPKMLSLKLE 345

```

Db      127 GSGTRKRLICQATGSPKQISVSWLRDQKVESGVLTKEVNEATKGAGPATPSISMLT 186
Qy      346 NKEAKYSKREKPYWVLANPEAGMOCILSDSGVLLSNITKVLPTWTPPCPAPEKSCD 405
Db      187 IRESD-----WL-----SQSLYTCRVDRHG-IFPKNVMSSECT-----TPSP----- 225
Qy      406 KHTCCELLGSPVFLPPKPKDTLMISRTPEVTCVVDVSDPEVKFNWYDVGEVHN 465
Db      226 -----GIVVFPIAPSFADT-FLSASRLICLVDTLTYG-SLNTSW-----ASHN 268
Qy      466 AKT-----KPERQVNSTYRVSVLTVLVHODWLNKGEYKCKVSNKALPAPIEKTISKAG 520
Db      269 GKALDTHMNTESHSPATFSGMEASVCAEDWESGQFTCTVHADLPPLKHTISKSRH 328
Qy      521 QPRE-FOVYTLPPSRBELT-KNOVSLTCLVKGFPSPDIAYEWSNGQP--ENNYKTPPY 576
Db      329 VAKHPAVVYLPAPREQLVLRBSATVCLVKGFPSPADVFOVMOQRQPLSSDCKVTSAA 388
Qy      577 LD--SDGSFELYSKLTVDKSRMOQGNVFGSCVMHEALHNHYTQKSLSPGLQDETCAE 634
Db      389 PEPQAPGLYFHTHTLTLYTEEDMNSGETFTTCVGHGHALPHWVTERIVDKS-----TEGEVG 443
Qy      635 AQDELDGLMTT 646
Db      444 AEEGFENLMTT 455

RESULT 40
MUCB_HUMAN STANDARD; PRT; 391 AA.
AC P04220;
DT 20-MAR-1987 (Rel. 04, Created)
DT 20-MAR-1987 (Rel. 04, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ig mu heavy chain disease protein (BOT).
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RX MEDLINE=6418166; PubMed=6425189;
RA Barnikol-Watanabe S., Mihaesco E., Mihaesco C., Barnikol H.U.,
RA Hilschmann N.,
RT "The primary structure of mu-chain-disease protein BOT. Peculiar
RT amino-acid sequence of the N-terminal 42 positions."
RL Hoppe-Seyler's Z. Physiol. Chem. 365:105-118(1984).
CC -1- MISCELLANEOUS: This protein has no V region homology or CH1
CC region.
DR PIR; A02163; MHUUT.
DR HSSP; P01857; 1PCL.
DR GO; GO:0005624; C:membrane fraction; NAS.
DR GO; GO:0003823; F:antigen binding; TAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-1like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IG1; 2.
DR PROSITE; PS00835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 3.
KW Immunoglobulin domain; Immunoglobulin C region.
FT DOMAIN 1 42 PRE-C-PART (NO V REGION HOMOLOGY).
FT DOMAIN 43 155 CH2.
FT DOMAIN 156 261 CH3.
FT DOMAIN 262 391 CH4.
SQ SEQUENCE 391 AA; 43057 MW; 9100843AF0CF021A CRC64;

Query Match 10.3%; Score 352.5; DB 1; Length 391;
Best Local Similarity 27.8%; Pred. No. 2,1e-15;
Matches 84; Conservative 61; Mismatches 110; Indels 47; Gaps 11;
332 GPSPTKMLSLKLENKAKYSKREKPYWVLANPEAGMOCILSDSGVLLSNITKVL--PT 389

```

```

Db      110 GPTTKVSTILTKESD-----WL-----GQSMFTCRVDHRG-LTFQQAASSKCGPD 155
Qy      390 WSPVPCPAPEKSCDKHTCCELLGSPVFLPPKPKDTLMISRTPEVTCVVDVSHD 449
Db      156 QDTAI-----RVFAIPPS-FASIFLTSTKTLCTLVDTLTYTD 191
Qy      450 PEVKFNWYDVGEVHNATKPREQVNSTYRVSVLTVLVHODWLNKGEYKCKVSNKALPA 509
Db      192 -SVTISWTRQDGEAVKTHNTNISESPNATFSAVGERSICEDDMDGERTCTVHTDPS 250
Qy      510 PIKTIISKAGQP-REPOVYTLPPSRBELT-KNOVSLTCLVKGFPSPDIAYEWSNGQP- 566
Db      251 PLKQITSRKGVALLRPDYLLPAPREQLNLRBSATITLTVGFSPADVFVQMOQRGQPL 310
Qy      567 -ENNYKTPPYLD--SDGSFELYSKLTVDKSRMOQGNVFGSCVMHEALHNHYTQKSLIS 623
Db      311 SPEKYVTSAPWEPQAPGRYFAHSILTVSEEMNTGETVTCVVAHEALPWRVTERIVDKS 370
Qy      624 PG 625
Db      371 TG 372

RESULT 41
MUC_MOUSE STANDARD; PRT; 455 AA.
AC P01872;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig mu chain C region secreted form.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RX MEDLINE=81076590; PubMed=6255422;
RA Kawakami T., Takahashi N., Honjo T.;
RT "Complete nucleotide sequence of mouse immunoglobulin mu gene and
RT comparison with other immunoglobulin heavy chain genes."
RL Nucleic Acids Res. 8:3933-3945(1980).
RN [2]
RX MEDLINE=82051295; PubMed=6795090;
RA Goldberg G.I., Vanin E.F., Zrolka A.M., Blattner F.R.;
RT "Sequence of the gene for the constant region of the mu chain of
RT Balb/c mouse immunoglobulin."
RL Gene 15:33-42(1981).
RN [3]
RX MEDLINE=79223904; PubMed=111247;
RA Kehrly M.R., Sibley C.H., Fuhrman J.S., Schilling J.W., Hood L.E.;
RT "Amino acid sequence of a mouse immunoglobulin mu chain."
RL Proc. Natl. Acad. Sci. U.S.A. 76:2932-2936(1979).
RN [5]
RX MEDLINE=83075344; PubMed=6816276;
RA Kehrly M.R., Fuhrman J.S., Schilling J.W., Rogers J., Sibley C.H.,
RA Hood L.E.;
RT "Complete amino acid sequence of a mouse mu chain: homology among
RT heavy chain constant region domains."
RL Biochemistry 21:5415-5424(1982).
CC -1- SUBCELLULAR LOCATION: Secreted (Probable).
CC -1- ALTERNATIVE PRODUCTS:

```

```

CC Event-Alternative splicing; Named isoforms=2;
CC Comment-During differentiation, B lymphocytes switch from
CC expression of isoform Membrane-bound to isoform Secreted;
CC Name=Secreted;
CC IsoId=I01872-1; Sequence=Displayed;
CC Name=Membrane-bound;
CC IsoId=I01873-1; Sequence=External;
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; V00827; -, NOT_ANNOTATED_CDS.
DR PIR; A02186; MHMS.
DR HSSP; P01857; IFC1.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; IgC1; 2.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KM Alternative splicing.
FT NON_TER 1
FT DOMAIN 1 105 CH1.
FT DOMAIN 106 217 CH2.
FT DOMAIN 218 324 CH3.
FT DOMAIN 325 436 CH4.
FT DOMAIN 437 455 C-TERMINAL SEGMENT OF SECRETED FORM.
FT DISULFID 14 14 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 28 89 BY SIMILARITY.
FT DISULFID 136 199 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 216 216 BY SIMILARITY.
FT DISULFID 246 305 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 293 293 BY SIMILARITY.
FT DISULFID 353 415 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 454 454 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .).
FT CARBOHYD 211 211 N-LINKED (GLCNAC. . .).
FT CARBOHYD 243 243 N-LINKED (GLCNAC. . .).
FT CARBOHYD 281 281 N-LINKED (GLCNAC. . .).
FT CARBOHYD 442 442 N-LINKED (GLCNAC. . .).
FT VARIANT 78 78 S -> N (IN MOPC 104E).
FT VARIANT 101 101 H -> Q (IN MOPC 104E).
FT VARIANT 226 226 T -> N (IN TEPIC183 AND MOPC 104E).
FT VARIANT 258 258 N -> T (IN TEPIC183).
FT VARIANT 258 258 N -> S (IN TEPIC183).
FT VARIANT 368 368 L -> K (IN TEPIC183 AND MOPC 104E).
SQ SEQUENCE 455 AA; 50101 MM; 4CBES7CB602F9B51 CRC64;
-----
Query Match 10.2%; Score 348; DB 1; Length 455;
Best Local Similarity 22.7%; Pred. No. 4.9e-15;
Matches 122; Conservative 77; Mismatches 179; Indels 160; Gaps 18;
-----
144 ESPPGSSPVQCRSP-RGKNIQGGKTLVSQLELQDSGTWTCTVLQNKVRFIDIVL 202
3 QSPFNVPFLVSCSPSLSKNLVAMGCLARDPLSTISTTN----- 43
203 AFQKASSIVKKEGQVEFSPFLAFTVEKLTGSGELMWQARASSKSWITFDLKNKEVS 262
44 -----YQNTTEVIG-----GIRTFPLRTGKXLA 68
263 VKRVTQDPKLM-----GKKLPHLTLPOALPOYAGSGULTALEAKTKGL 308
69 TSQVLLSPKSLTEGSDYLCKIHYGKGRDLHVPF-----AVAEW 110
309 HOEVNLVV-----MRAITQLQKNTLCEVWGPSTPKMLSLTKLENKAKVSKREKPVWVNL 362

```

```

DB 111 NPVNVFVPPRDESGAPRKRSLICEATNFTPKRITVSMKQGLVESGETTDPVTIEN 170
QY 363 ---PEA-----GMWQCLSDSGVLLSENIKVLPTWSTPVPCPAPEP 401
DB 171 KGSITPQYKVIISTLTITSEIDMLNLYTCRDYHGLFLK----- 210
QY 402 KSCDKTHTCPELLGSPS---VFLPPPKVDTLMISTRPEVTCVVDVSHEDPEVKRWY 457
DB 211 ---NVSTTC---AASPSTDLITFTIPSPAD-IFLSASANTLCVSLMAYE-TLNIWSA 262
QY 458 VDVGVFNNAKTPREEOYNSTYRVSVLYTLHODMLNGEKYCKVSKALPAPIEKTISK 517
DB 263 SOSGEPELETIKIMESHNGTFSAKVASVCEEDMNRKEVCTVTRDLPSPQKFKISK 322
QY 518 AKGQPRE-----POVYTLPPSRDEL-T-KNOVSLTCLVKGFPSPDIIVEMESNGO--PENN 569
DB 323 ---PNEVHKHPPVAVLLPPAREQNLNRESATVTCVKGSPADISVQMLQGLLPQEK 378
QY 570 YKTPPYLD--SDGSFPLYSKLTVDKSRWQGNVFCSCVMEHALNHYTKSLSPG 625
DB 379 YVTSAPMPEGAPGFYFTHTSLITVEEENSGEYTCVVGHEALPHLVTERTVDKSTG 436
-----
RESULT 42
MUC_SUNMU STANDARD; PRT; 457 AA.
ID MUC_SUNMU
AC P20768;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Ig mu chain C region.
OS Suncus murinus (House shrew) (Musk shrew).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Insectivora; Soricidae; Crocodyrniae; Suncus.
OX NCBI_Taxid=9378;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Liver;
RX MEDLINE=69232144; PubMed=2497033.
RA Ishiguro H., Ichihara Y., Nankawa T., Nagatsu T., Kurosawa Y.;
RT "Nucleotide sequence of Suncus murinus immunoglobulin mu gene and
RT comparison with mouse and human mu genes";
RL FEBS Lett. 247:317-322(1989).
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
DR EMBL; X13920; CAA3113.1; ALT_INIT.
DR PIR; S03961; S03961.
DR HSSP; P01842; TRAB.
DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 4.
DR SMART; SM00407; IgC1; 3.
DR PROSITE; PS00835; IG_LIKE; 4.
DR PROSITE; PS00290; IG_MHC; 3.
KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.
FT NON_TER 1
FT DOMAIN 1 105 CH1.
FT DOMAIN 106 220 CH2.
FT DOMAIN 221 326 CH3.
FT DOMAIN 327 457 CH4.
FT DISULFID 13 13 INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT DISULFID 27 89 BY SIMILARITY.
FT DISULFID 136 200 BY SIMILARITY.
FT DISULFID 217 217 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT DISULFID 248 307 BY SIMILARITY.

```

FT DISULFID 295 295 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).  
 FT DISULFID 355 417 BY SIMILARITY.  
 FT DISULFID 456 456 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).  
 FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 113 113 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 276 276 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 283 283 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 444 444 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 457 AA; 50074 MW; 56C8C6BDA4462B9 CRC64;

Query Match 10.1%; Score 346; DB 1; Length 457;  
 Best Local Similarity 24.3%; Pred. No. 6.7e-15;  
 Matches 112; Conservative 89; Mismatches 156; Indels 104; Gaps 20;

QY 204 FQKASSIVYKKEGEQVFFPLAFVYKLTGSGELMWQABR-ASSSKWTFDLKN- 258  
 43 FKSSSI-----SSQNIYFPEVFTGKVMATSQVLLPSTALQSTDDYITCKTKTGE 97  
 QY 259 KEYSYKRVTDPKLQMGKLPPLHLTL-POALPOYAGSGNLTALAKTKGKLMQEVNLVY 316  
 98 KEKKV-----ELQVPELPPNVSIFVP--PRNSFSGN-----HRTSQLL 135  
 QY 317 MRATQLOKNTCEVW-----GPTSPKMLSLKENKAKVS 352  
 136 CQASGSPRTVWMLQGEFVQPSLVSTSAVLEPKGSGPTFRVLSLTITENE----- 191  
 QY 353 KREKPVWLNPEAGMOCLLSDSGQVLLSENIKVLPTWSTPVPCEPAEPRKSC--DKTHTC 410  
 192 -----WLSQRE--FTQALHKG-LTFQKNVSV-----CMGDDTST- 224  
 QY 411 PELLGSPVFLFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNAKTKP 470  
 225 -----GISVELLPPTFAN-IFLTQSAQLTCLVGLATYD-SLDISWRONGEALQTHVNI 277  
 QY 471 REQVNTYRVVSVLTYLHODMLNGEKYCKKVSINKALPAIETKTSKAKQPRE-POVYT 529  
 278 SEHNPSTFTAKGHASVCREWESGEKFTCTVQHSDLPSLKOSLSRPKVANDPSPSVFV 337  
 QY 530 LPPSRDEL-TKNQVSLTCLVKGFPSPDAVWESNGOP--ENNYKTPPVLD--SDGSFF 584  
 338 LPPAQEDELKLRASASTICLVKSPRPVFWQMHGQRPVDPKHYVSNPFRPNQNPPLYF 397  
 QY 585 LYKSLTYDKSRWQGNVFCSCVMHEALHNHTQKSLSLSG 625  
 398 VHSILTVSEKDWSSGESFCVGHGHALPLSTVERKADVKTSG 438  
 DB

RESULT 43  
 MUC\_RABIT  
 ID MUC\_RABIT STANDARD; PRT; 458 AA.  
 AC P03988;  
 DT 23-OCT-1986 (Rel. 02, Created)  
 DT 23-OCT-1986 (Rel. 02, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE 1g mu chain C region secreted form.  
 OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 NC NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A. (A2 ALLOTYPED).  
 RX MEDLINE=84088930; PubMed=6418803;  
 RA Bernstein K.E., Alexander C.B., Reddy E.P., Mage R.G.;  
 RT "Complete sequence of a cloned cDNA encoding rabbit secreted mu-chain  
 of vha2 allotype: comparisons with vha1 and membrane mu sequences.";  
 RL J. Immunol. 132:490-499(1984).  
 CC -1- SUBCELLULAR LOCATION: Secreted (Probable).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Comment=During differentiation, B lymphocytes switch from  
 CC expression of isoform Membrane-bound to isoform Secreted;  
 CC Name=Secreted;

CC IsoId=P03988-1; Sequence=Displayed;  
 CC Name=Membrane-bound;  
 CC IsoId=P04221-1; Sequence=External;  
 -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

CC EMBL; K01357; NOT\_ANNOTATED\_CDS.  
 CC PIR; A02164; MHRB.  
 CC HSSP; P01842; 7FAB.  
 DR Interpro; IPR007110; Ig-like.  
 DR Interpro; IPR003597; Ig\_c1.  
 DR Interpro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig\_4.  
 DR SMART; SM00407; IgC1; 2.  
 DR PROSITE; PS00835; IG\_LIKE; 4.  
 DR PROSITE; PS00290; IG\_MHC; 3.  
 KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;  
 KW Alternative splicing.  
 FT NON TER 1  
 FT DOMAIN 1 106 CH1.  
 FT DOMAIN 107 222 CH2.  
 FT DOMAIN 223 327 CH3.  
 FT DISULFID 14 14 CH4.  
 FT DISULFID 28 90 BY SIMILARITY.  
 FT DISULFID 137 200 BY SIMILARITY.  
 FT DISULFID 219 219 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).  
 FT DISULFID 249 306 BY SIMILARITY.  
 FT DISULFID 296 368 INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).  
 FT DISULFID 356 418 BY SIMILARITY.  
 FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 114 114 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 212 212 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 261 261 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 277 277 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 284 284 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 445 445 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 458 AA; 49897 MW; F33BD6A3B2B84864 CRC64;

Query Match 10.1%; Score 346; DB 1; Length 458;  
 Best Local Similarity 24.8%; Pred. No. 6.7e-15;  
 Matches 117; Conservative 78; Mismatches 168; Indels 108; Gaps 20;

QY 223 FPL-----AFVYKLTGSGELMWQABRASSSKWTFDLKNVEVYKRVTDPKLQMGKK 277  
 9 YPLVSCGALTDGNLWAGCLARDFLPSSVTSW-SFK-NSBSISRTVTRTFVVRGDK 66  
 QY 278 -----LPLHLTLPOAL-----POYA 292  
 67 YNATSOVLVPSKDVLOGTEBYLVCKQHSNRLDRAVSPVSELPNNVSFTRPPDSFS 126  
 QY 293 GSGNLTALAEK-TGKLHDEVNLVVRATQ-----LQKNTCEVWGPTSPKMLSLKLE 345  
 127 GSGTRSRILICQATGSPKQISVWLRDQKVESGLVTRPVEAFETGAGPATFSISMLT 186  
 DB 346 NKEAKVSKREKRPVWLNPEAGMOCLLSDSGQVLLSENIKVLPTWSTPVPCEPAEPRKSCD 405  
 187 ITESD-----WL-----SQSLVTCVDRHG-IFPKNVVMSSECT-----TPSP- 225  
 QY 406 KHNHCCELGGSPVFLFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKFNMYVDGVEVNH 465  
 226 -----GLOVFIAPSPDT-FLSKSARILICVTDITTYG-SLNSW-----ASHN 268  
 DB 466 AKT-----KREQVNTYRVVSVLTYLHODMLNGEKYCKKVSINKALPAIETKTSKAKG 520  
 269 GKALDTHNMITEHPRATISAMGEASVCAEDMESGQFTCTVHADLPFLPKHTISKRSR 328

Oy	521	CPRE--POVYTLPPSPRELT-KNOVSLTCLVKGPSPDIDIVEMSNOP--ENNYKTPPV	576
Db	329	VAKHPAYVYVLEPAREQLVRESATYTCVLKGFSPADVGVQMOQRQPLSSDKYVTSAPA	388
Oy	577	LD--SDGSFPLYSKLTVDKSRWQOQNVBSCSVMEHLAHNYTOKSLSLSPG	625
Db	389	PEPQAPGLYFTHTSTLTVEEDNNSGTEFLFCVVGHEHLLPHAVYERTYDKSG	439

## RESULT 44

ID	MUC MESAU	STANDARD:	PRT:	454 AA.
AC	P06337.			
DT	01-JAN-1988	(Rel. 06, Created)		
DT	01-FEB-1996	(Rel. 33, Last sequence update)		
DT	15-JUL-1999	(Rel. 38, Last annotation update)		
DE	Ig mu chain C region.			
OS	Mesocricetus auratus (Golden hamster).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;			
CC	Mesocricetus.			
OX	NCBI_TaxID=10036;			
RN	{1}			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE=85297761; PubMed=2994005;			
RA	McGuire K.L., Duncan W.R., Tucker P.W.;			
RT	"Phylogenetic conservation of immunoglobulin heavy chains: direct			
RL	comparison of hamster and mouse Cmu genes."			
CC	Nucleic Acids Res. 13:5611-5628(1985).			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see <a href="http://www.isb-sdb.ch/announce/">http://www.isb-sdb.ch/announce/</a>			
CC	or send an email to <a href="mailto:license@sdb-sdb.ch">license@sdb-sdb.ch</a> ).			
CC	-----			
DR	EMBL; X02804; CAA26574.1; -.			
DR	HSSP; P01854; 1IGE.			
DR	InterPro; IPR007110; Ig-1Ike.			
DR	InterPro; IPR003597; Ig_C1.			
DR	InterPro; IPR003006; Ig_MHC.			
DR	Pfam; PF00047; Ig; 4.			
DR	SMART; SM00407; IGc1; 2.			
DR	PROSITE; PS00835; IG_LIKE; 4.			
DR	PROSITE; PS00290; IG_MHC; 3.			
KW	Immunoglobulin domain; Immunoglobulin C region; Glycoprotein.			
FT	NON TER	1		
FT	DOMAIN	1	105	CH1.
FT	DOMAIN	106	218	CH2.
FT	DOMAIN	219	324	CH3.
FT	DOMAIN	325	454	CH4.
FT	DISULFID	13	13	INTERCHAIN (WITH LIGHT CHAIN) (PROBABLE).
FT	DISULFID	27	88	BY SIMILARITY.
FT	DISULFID	135	198	BY SIMILARITY.
FT	DISULFID	215	205	INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT	DISULFID	245	304	BY SIMILARITY.
FT	DISULFID	292	292	INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT	DISULFID	352	414	BY SIMILARITY.
FT	DISULFID	453	453	INTERCHAIN (WITH HEAVY CHAIN) (PROBABLE).
FT	CARBOHYD	45		N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	112	112	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	192	192	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	210	210	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	238	238	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	257	257	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	280	280	N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD	441	441	N-LINKED (GLCNAC. . .) (POTENTIAL).
SO	SEQUENCE	454 AA;	50196 MW;	4DA2134612BB1469 CXC64;

Query Match 10.0%; Score 342.5; DB 1; Length 454;

Best Local Similarity 25.7%; Pred. No. 1,1e-14;

## RESULT 45

ID	MUC CANPA	STANDARD;	PRT;	450 AA.
AC	P01874;			
DT	21-JUL-1986	(Rel. 01, Created)		
DT	21-JUL-1986	(Rel. 01, Last sequence update)		
DT	15-JUL-1999	(Rel. 38, Last annotation update)		
DE	Ig mu chain C region.			
OS	Canis familiaris (dog).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.			
XX	NCBI_Taxid=9615;			
RN	[1]			
RP	SEQUENCE OF 1-177 (MOO).			
RK	MEDLINE=80077682; PubMed=117299;			
RA	McClumber L.J., Capra J.D.;			
RT	"The complete amino-acid sequence of a canine mu chain.";			
RL	Mol. Immunol. 16:565-570 (1979).			
RN	[2]			
RP	SEQUENCE OF 178-450 (MOO).			
RK	MEDLINE=78180587; PubMed=653360;			
RA	Wasserman R.L., Capra J.D.;			
RT	"Amino acid sequence of the Fc region of a canine immunoglobulin M:			
RL	Science 200:1159-1161 (1978).			
DR	PIR, A93131; MHDG.			
DR	HSSP; P01857; 1FCL.			
DR	InterPro; IPR007110; IG_1like.			
DR	InterPro; IPR003597; IG_c1.			
DR	InterPro; IPR003006; IG_MHC.			
DR	Pfam; PF00047; Ig; 4.			
DR	SMART; SM00407; IGC1; 2.			
DR	PROSITE; PS50835; IG_LIKE; 4.			
DR	PROSITE; PS00290; IGMHC; 3.			
FT	Immunoglobulin domain; Immunoglobulin C region.			
FT	NON TER			
SEQUENCE	450 AA;	48895 MW;	90460DAD9D1012FSD	CRC64;

```

Query Match      10.0%; Score 340; DB 1; Length 450;
Best Local Similarity 25.9%; Pred. No. 1.6e-14;
Matches 100; Conservative 76; Mismatches 142; Indels 68; Gaps 15;

QY 278 LPLHLTLPOOL-----POYAGSGN--LTLALEAKTKGLHOEYNLVMBATOLQKNLT-- 327
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 102 LPLVMTLPPREVSGFIPPRDAFGBPRKSQULICASGFSRQVWSL--RDGKQIESGVTIN 159
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 328 -----CEVWGPTSPKMLSLKLENKAKVSKREKPVVTLNPEAGMOCILSDSGVLLAS 382
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 160 EVZAZAKZSGPTTYKYKTSMLTI-----QEDMWL--SQSVPTCKVEHNG-LTFQO 205
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 383 NIKVLPMTSTPVCAPAPRPSCKHTHCPELLGSPVFLPPPKXDTLMISRPEVTCV 442
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 206 NASSMCTSDQPV-----GISIFITTPPS-FASINTYSAKLSCLV 243
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 443 VDVSHDEPEVKFNMYVDGVEVHNAKTPREEOYNSTYRVSVLTVLHODMLNKEVCKV 502
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 244 TDLATVD-SVTLSWTEENGALTKHTNISHSHNGFMSAMGEATVCVEEMESGEQFTCTV 302
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 503 SNKALPAPIEKTISKAKG-QPREPOVYTLPPSRDEL-TKNQVSLTLVKGFPYSDIAVEM 560
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 303 THTDLPVSLVKQITSRKGVAVHMPVSYVLPSPREQLDRESATLSCLVGYSPDPVQVM 362
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 561 ESMGO--PENNYKTPPVLD--SDGSFPLYSKLTVDKSRHOGQNVSCSVMHEALHNHYT 616
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 363 VQKGQVPPDQSVYTSAPMEBPQAPGLYFAHSILTVSEEMNAGETTYTCVAHESLPRVT 422
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 617 OKSLSLSPG-----LQIDETCAE 634
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 423 ERSVDKSTGKPTLYNYSVLSDTAGZ 448
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 46
ALC_RABIT STANDARD; PRT; 299 AA.
ID ALC_RABIT
AC P01879;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig alpha chain C region (Fragment).
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_Taxid=9986;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=84144059; PubMed=6322114;
RA Knight K.L., Martens C.L., Stoklosa C.M., Schneiderman R.D.;
RT "Genes encoding alpha-heavy chains of rabbit IgA: characterization of
   RT CDNA encoding IgA-g subclass alpha-chains."
RL Nucleic Acids Res. 12:1657-1670(1984).
CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body
   CC secretions. It may serve both to defend against local infection
   CC and to prevent access of foreign antigens to the general
   CC immunologic system.
CC -1- MISCELLANEOUS: This immunoglobulin belongs to the IgA-G subclass.
   CC It was isolated from a rabbit homozygous FOR A2, N80, DE12, 15,
   CC F71, G75 heavy chain haplotype.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X00353; CAA25100.1; -.
CC PIR; A02174; AHRB.
CC HSSP; P01857; 1FCL.
CC InterPro; IPR007110; Ig-1like.

```

```

DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1.2.
DR SMART; SM00407; IgC1.2.
DR PROSITE; PS50835; IG_LIKE.2.
DR PROSITE; PS00290; IG_MHC.2.
KW Immunoglobulin domain; Immunoglobulin C region; Repeat.
FT NON_TER 1
FT DOMAIN 1 71 167 IG-LIKE 1.
FT DOMAIN 174 276 IG-LIKE 2.
SQ SEQUENCE 299 AA; 32256 MW; 2512FE3F62E9A223 CRC64;

Query Match      9.0%; Score 306; DB 1; Length 299;
Best Local Similarity 32.2%; Pred. No. 1.3e-12;
Matches 88; Conservative 42; Mismatches 115; Indels 28; Gaps 13;

QY 369 QCLSDSGVLLSNIKVLPTWSTPV--CPAPRPSCKHTHCP-----ELIGG--PSV 419
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 20 QCLGQKSAACHVEYNSVINESLPVPFPDCCPA---NSC---CTCPSSSRNLISGQPSL 73
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 420 FLPPPKDITLMISRPEVTCVVDVSHDEPEVKFNMYVDGVEVHNAKTPREEOYNSTY 479
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 74 SLQRPDLGD-LLLGDASLTCTLSGLKNEQAV-FTW--EPTNGNEPVQORQRDISGCT 129
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 480 RVVSVTLTVLHODMLNKEVCKVSNKALPA-DIEKTISKAKQPREPOVYTLPPSRDEL 538
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 130 SVSVSLPSSAEFTWKARTERTCTVTHTPEIDSGSLTATISGVTP--POVHLPPREBELA 187
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 539 KN-QVSLTCLVKGFPYPSDIAVEMESNGQ--PENNY--KTPPVLDSDGSFPLYSKLTVD 592
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 188 LNEQVTLTCLTVAGFSPKDVLVSMRHOQGEVPEPDSFLVWKMSPSSQDKATVATISLRLVP 247
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 593 KSRMOGNVFCSSVMHEALHNHYTOKSLSLSPG 625
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 248 AEDMNGDTYSCVGHGLAEHFTKTDRLAG 280
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 47
HVCN_HETFR STANDARD; PRT; 461 AA.
ID HVCN_HETFR
AC P23088;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig heavy chain C region, membrane-bound form (Clone 3050).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_Taxid=7792;
RN (1)
RP SEQUENCE FROM N.A.
RX TISSUE=Spleen;
RX MEDLINE=88328985; PubMed=3138109;
RA Kokubu F., Hinde K., Litman R., Shambloet M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
   RT constant region genes in a phylogenetically primitive vertebrate."
RL EMBO J. 7:1979-1988(1988).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; X07781; CAA30614.1; ALT_SEQ.
CC PIR; S01854; HYRKO.
CC HSSP; P01842; 7FAB.
CC InterPro; IPR007110; Ig-1like.
CC InterPro; IPR003597; Ig_c1.
CC InterPro; IPR003006; Ig_MHC.

```



```

Db      327  USAORFLSLTCLVGRGFPRFIFVKTVDKSVNPGNYKNTVMAENDNSSFYISLLSLIA 386
Oy      593  KSRWQGVNFCSCVWHENLHHHYTKSLSLSPG 625
Db      387  AEEWASGASVSCVGHENLPIKINRTVNKSSG 419

RESULT 49
MUC_CHICK
ID MUC_CHICK STANDARD: PRT; 446 AA.
AC P01875;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Ig mu chain C region.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Gallus.
CX NCBI_TaxID=9031;
RN (1)
RN SEQUENCE OF 80-446 FROM N.A.
RX MEDLINE=83299221; PubMed=6310496;
RA Dahan A., Reynaud C.-A., Weill J.-C.;
RT "Nucleotide sequence of the constant region of a chicken mu heavy
RL chain immunoglobulin mRNA."
RL Nucleic Acids Res. 11:5381-5389(1983).
RN (2)
RN SEQUENCE OF 1-79 FROM N.A.
RA Weill J.-C.;
RL Submitted (MAR-1986) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL, X01613, CAA25762.1; -.
DR PIR, A02170; MHCH.
DR HSSP, P01857; 1RC1.
DR InterPro, IPR007110; Ig-1-like.
DR InterPro, IPR003597; Ig_C1.
DR InterPro, IPR003006; Ig_MHC.
DR Pfam, PF00047; Ig: 3.
DR SMART, SM00407; IGc1, 1.
DR PROSITE, PS50835; IG_LIKE; 4.
DR PROSITE, PS00290; IG_MHC; 3.
DR Immunoglobulin domain; Immunoglobulin C region.
KW NON_TER
FT DOMAIN 1 105 CH1.
FT DOMAIN 106 209 CH2.
FT DOMAIN 210 316 CH3.
FT DOMAIN 317 427 CH4.
FT DOMAIN 428 446 C-TERMINAL REGION.
FT DOMAIN 16 16 INTERCHAIN (WITH A LIGHT CHAIN)
FT (BY SIMILARITY).
FT (BY SIMILARITY).
FT DISULFID 27 85 BY SIMILARITY.
FT DISULFID 135 190 BY SIMILARITY.
FT DISULFID 237 296 BY SIMILARITY.
FT DISULFID 284 284 INTERCHAIN (WITH A HEAVY CHAIN)
FT (BY SIMILARITY).
FT (BY SIMILARITY).
FT DISULFID 344 406 INTERCHAIN (WITH A HEAVY CHAIN)
FT DISULFID 445 445 (BY SIMILARITY).
FT CARBOHYD 45 45 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 46 46 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 73 73 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 130 130 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 198 198 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 382 382 N-LINKED (GLCNAC. . .) (POTENTIAL).

```

```

FT CARBOHYD 433 433 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 446 AA; 48173 MW; 3CB0CE108949BD1 CRC64;

Query Match 8.8%; Score 300; DB 1; Length 446;
Best Local Similarity 27.9%; Pred. No. 5,5e-12;
Matches 116; Conservative 61; Mismatches 179; Indels 60; Gaps 21;

OY SSKSWITFEDLKNKEVSVKRTODPKLQMGKKLPLHLTLPOLPOYAGSGNLTLLEAKTG 306
DB 35 SSIAITFWFDNNSSVSGMDVI--PKYISG-----PPYRVSRR--IQMNOSEG 77
OY KLHGEVVLVVMRATOLQKNTLCEVWGP--TSPK--LMLSLKLENK----- 348
DB 78 KEKQPPR---CRAHPRGNVEVSNMNGPIPTENGPIPLFTMHPPRSNDEEGPPRNASTL 134
OY 349 AKVSKREKP--VWVLN--PEAGMOCCLSDSGQVLIENIKQLPT--WST--PVPCAP- 399
DB 135 CQTRGRRRPTEVMTYKNGSPVAAATATATTATVGEVVAESRISVTESEMDTGATFSCVVEG 194
OY 400 EPKSCDTHHC---PELLGSPVFLFPPRKQDMLMTSRTEGVNCCVVVDVSHEDP-EVKFN 455
DB 195 EMRTSSRMECGLEPVPVQODIALRVITPSPVD-IFISKSTLTLGRVSNMVAQGLEVSW- 252
OY 456 WYVDGVVHNAAKTPREEQYNSTYRVVSVLTVLHODVLNKEYCKVSNALPAPIEKT1 515
DB 253 WKKGGLLEFRLALRG-RVLQNGSLTYVDGVAITVCASEMDGGDGYVCKKNHDDLPEPMEM 311
OY 516 SKAK-GQPREPQVYTLRPSRDELTKNQ-VSLTCLVKGFPYSDIAVWESNQ--PENNYK 571
DB 312 RKTASNAARPPSVYVFPPTPEQLNGNQRSLSVTCMAQGFNPPLFVFMWRNGEPLPQSQSV 371
OY 572 TTPPVLDLS--DGSFYLKSLTVDKSRNQGNVSSCSYMHALNHNHYOKSLISPG 625
DB 372 TSAEMAENPENESYVAVSVLGVGAEEWAGNVYTCVLGHEALPLQLAKQSVDRASG 427

RESULT 50
HCVC2_HETFR STANDARD; PRT; 438 AA.
AC P23085;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig heavy chain C region (Clone 12022) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontoidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;

[1]
RN SEQUENCE FROM N.A.
RP MEDLINE=88328985; PubMed=3138109;
RA Koriibu F., Hinds K., Litman R., Shamloct M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
RL EMBO J. 7:1979-1988(1988).
-- SIMILARITY: Contains 4 immunoglobulin-like domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; X07784; CA930617.1; -
DR PIR; S00980; HVKRC2.
DR HSSP; P01842; 7FAB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig cl.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 4.

```



DR	SMART; SMO0407; IGc1; 2.
DR	PROSITE; PS50835; IG_LIKE; 4.
DR	PROSITE; P800290; IG_Mc; 3.
KW	Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT	NON_TER 1 1
FT	DOMAIN 1 101 IG-LIKE 1.
FT	DOMAIN 108 202 IG-LIKE 2.
FT	DOMAIN 213 305 IG-LIKE 3.
FT	DOMAIN 315 415 IG-LIKE 4.
FT	CARBOHYD 166 166 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 200 200 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 245 245 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 275 275 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 374 374 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 411 411 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 415 415 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT	CARBOHYD 425 425 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ	SEQUENCE 438 AA; 47904 MW; 6667A9C5BBF7AA7B CRC64;
Query Match	8.7%; Score 297.5; DB 1; Length 438;
Beat Local Similarity	25.2%; Pred. No. 7.7e-12;
Matches 113; Conservative 72; Mismatches 167; Indels 97; Gaps 19;	
OY	208 SSIVYKKGEVSEFSPLAFV-----EKLNGSELMMQBRASSKSW 251
DB	37 TSITMKDKEPITTGKTIYPVLANKGYTRSSQLTTESEVGSKTYCEVRGHS--LW 94
OY	252 ITFDLNKNREYSKYRVTOPKLOMGKKLPRLHLTPQALPOYAGSGMLTALAETGTLHOE 311
DB	95 I-----KEI-----LOCK---GDIVPEPTVILTGSSEBEISRFPATVLCIIIDPHPS 139
OY	312 VNLVVMRAIQ----LQNNLTCEWGAPTSFKMLSLKLENKEAKYKREKPYVNLNPAEG 366
DB	140 ITVSMLKGGQPMDSGFVITSPTCEVANGNFSASTRLTVPAGE-----WFSNT--- 184
OY	367 MMQCILSDSGOYLLESNIKVLPWTSPNPSCAPAREPKSDCKHTTCPELLGSPSVLFPPKP 426
DB	185 YTTCCVAHQ-EYTOSRNIT-----GSQVPCS-----IGDPYIKLLPISI 222
OY	427 KDTLMISRTPEVTVCVAVVDSHEDPEVKFMVVYDGVEVNNAKTKEPREEQ-----NSTYRV 481
DB	223 EGVV-LBATVTLITCV---VSNAFYGVNSM-----TQEKKPLKSELAVPGHSDSIV 270
OY	482 VSULTLVLDHWLNKRYEKCKYSNKALLPAIEETISKANGQP-REPQV-YTLPPSRDELTK 539
DB	271 ISTVNIISTQAWLSGAEFYCVCVSHODLPFLRASIHKEEVKDRLRPBFVSVTLPPAEADVSAQ 330
OY	540 NOVSTICLVKGYRPPDIADVENMSNOQPER-NYKTTTPVLDSDG-SFLYXSLUTYDXSRW 596
DB	331 RFLSLTCLVRCGSPREFIFIKMTVNDKSVPNGYKYTEVAENDNRSPFIYSLISAEEW 390
OY	597 QCGNVFSCSYMHMALHNHYOKSLSPG 625
DB	391 ASGASYSCVCGHEALIPKINRTVKKSSG 419
RESULT 51	
ID	ALC2_HUMAN STANDARD; PRT; 340 AA.
AC	P01877.
DT	21-JUL-1986 (Rel. 01, Created)
DT	01-FEB-1991 (Rel. 17, Last sequence update)
DT	15-MAR-2004 (Rel. 43, Last annotation update)
DE	IG alpha-2 chain C region.
GN	IGHA2
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
OX	NCBI_Taxid=9606;
RN	[1]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=84130179; PubMed=6421489;
RA	Flanagan J.G., Lefranc M.-P., Rablitt T.H.;

RT	1	"Mechanisms of divergence and convergence of the human immunoglobulin
RT	2	alpha 1 and alpha 2 constant region gene sequences.";
RL	36:681-688(1984).	
RN	[2]	
RP	SEQUENCE (BUT).	
RA	MEDLINE=78137069; PubMed=416441;	
RX	Torano A., Putnam F.W.;	
RT	"Complete amino acid sequence of the alpha 2 heavy chain of a human	
RT	IgA2 immunoglobulin of the A2m (2) allotype.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 75:966-969(1978).	
RN	[3]	
RP	SEQUENCE (MYELOMA PROTEIN LAN).	
RX	MEDLINE=79180140; PubMed=286295;	
RA	Tsuzukida Y., Wang C.-C., Putnam F.W.;	
RT	"Structure of the A2m(1) allotype of human IgA -- a recombinant	
RT	molecule.";	
RL	Proc. Natl. Acad. Sci. U.S.A. 76:1104-1108(1979).	
RN	[4]	
RP	REVIEW	
RX	MEDLINE=91054387; PubMed=2241915;	
RA	Kerr M.A.;	
RT	"The structure and function of human IgA.";	
RL	Biochem. J. 271:285-296(1990).	
CC	-1- FUNCTION. Ig alpha is the major immunoglobulin class in body	
CC	secretions. It may serve both to defend against local infection	
CC	and to prevent access of foreign antigens to the general	
CC	immunologic system.	
CC	-1- SUBUNIT. Monomeric or polymeric.	
CC	-1- MISCELLANEOUS. The sequence of the A2m(1) allotype is shown.	
CC	-1- SIMILARITY. Contains 3 immunoglobulin-like domains.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb.ch/announce/or send an email to license@isb-sib.ch">http://www.isb.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
DR	EMBL; J00221; AAB59396.1; ALT_INIT.	
DR	PIR; B22360; B22360.	
DR	HSSP; P01810; 2FBJ	
DR	Genew; HGNC:5479; IGHA2.	
DR	MLN; I47000; -	
DR	GO; GO:0005624; C:membrane fraction; NAS.	
DR	GO; GO:0003823; F:antigen binding; TAS.	
DR	GO; GO:0006955; P:immune response; NAS.	
DR	InterPro; IPR007110; Ig-like.	
DR	InterPro; IPR003597; Ig_ct.	
DR	InterPro; IPR003006; Ig_MHC.	
DR	Pfam; PF00047; Ig; 3.	
DR	SMART; SM00407; Ig_Like; 2.	
DR	PROSITE; PS50835; IG_LIKE; 3.	
DR	PROSITE; PS00290; IG_MHC; 2.	
DR	Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.	
RW	NON_TER	
FT	1	
FT	DOMAIN	6 98
FT	DOMAIN	112 207
FT	DOMAIN	215 317
FT	DISULFID	26 85
FT	DISULFID	101 101
FT	DISULFID	109 109
FT	DISULFID	110 167
FT	DISULFID	134 191
FT	DISULFID	169 169
FT	DISULFID	179 179
FT	DISULFID	237 300
FT	DISULFID	339 339
FT	CARBOHYD	47 47
FT	CARBOHYD	92 92
FT	CARBOHYD	131 131
FT	CARBOHYD	205 205
FT	DISULFID	237 300
FT	DISULFID	339 339
FT	CARBOHYD	47 47
FT	CARBOHYD	92 92
FT	CARBOHYD	131 131
FT	CARBOHYD	205 205
FT	DISULFID	237 300
FT	DISULFID	339 339
FT	CARBOHYD	47 47
FT	CARBOHYD	92 92
FT	CARBOHYD	131 131
FT	CARBOHYD	205 205
FT	DISULFID	237 300
FT	DISULFID	339 339
FT	CARBOHYD	47 47
FT	CARBOHYD	92 92
FT	CARBOHYD	131 131
FT	CARBOHYD	205 205
FT	DISULFID	237 300
FT	DISULFID	339 339
FT	CARBOHYD	47 47
FT	CARBOHYD	92 92
FT	CARBOHYD	131 131
FT	CARBOHYD	205 205
FT	DISULFID	237 300
FT	DISULFID	339 339
FT	CARBOHYD	47 47
FT	CARBOHYD	92 92
FT	CARBOHYD	131 131
FT	CARBOHYD	205 205
FT	DISULFID	237 300
FT	DISULFID	339 339
FT	CARBOHYD	47 47
FT	CARBOHYD	92 92
FT	CARBOHYD	131 131
FT	CARBOHYD	205 205
FT	DISULFID	237 300
FT	DISULFID	339 339
FT	CARBOHYD	47 47
FT	CARBOHYD	92 92
FT	CARBOHYD	131 131
FT	CARBOHYD	205 205
FT	DISULFID	237 300
FT	DISULFID	339 339
FT	CARBOHYD	47 47
FT	CARBOHYD	92 92
FT	CARBOHYD	131 131
FT	CARBOHYD	205 205
FT	DISULFID	237 300
FT	DISULFID	339 339
FT	CARBOHYD	47 47
FT	CARBOHYD	92 92
FT	CARBOHYD	131 131
FT	CARBOHYD	205

```

CC CARBOHYD 327 327 N-LINKED (GLCNAC. .).
FT VARIANT 93 93 P -> S (IN A2M(2) ALLOTYPE).
FT VARIANT 102 102 /FTID=VAR_003879.
FT VARIANT 102 102 P -> R (IN A2M(2) ALLOTYPE).
FT VARIANT 279 279 /FTID=VAR_003880.
FT VARIANT 279 279 F -> Y (IN A2M(2) ALLOTYPE).
FT VARIANT 296 296 /FTID=VAR_003881.
FT VARIANT 296 296 D -> E (IN A2M(2) ALLOTYPE).
FT VARIANT 326 326 /FTID=VAR_003882.
FT VARIANT 326 326 V -> I (IN A2M(2) ALLOTYPE).
FT VARIANT 335 335 /FTID=VAR_003883.
FT VARIANT 335 335 V -> A (IN A2M(2) ALLOTYPE).
SQ SEQUENCE 340 AA; 36508 MM; 989227007563276 CRC64;

Query Match
Best Local Similarity 28.1%; Pred. No. 1e-11;
Matches 99; Conservative 46; Mismatches 140; Indels 67; Gaps 15;

QY 333 PTSPKLM-LSLKLENKAKSKR-----EKPPVWLNPEAGM-----WQCLSDSQ 377
DB 3 PTSPKVPPLSLDSTPDGNNVAVACLVQGFPPPLSLVSESGQNTARNPPSPDASGD 62
QY 378 VLLESNIKVLPTSTP-----VPCPAPPKSCDKTHTCPELIGP 417
DB 63 LYTTSSQTLTPATQCDGKSVTGHVKNYTPSDVTPVPCPPPPPC-----C-----HP 112
QY 418 SVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKFNNYVVDGEVHNAAKTRPEEQYNS 477
DB 113 RLSLHRLPALED-LLLGSEANLTCTLTGL-RDASGAFETWTPSSGK--SAVQGPREDLGG 168
QY 478 TYRNVSVLTLDQMDLNGKRYKCKVSKALPAIEKTIISKAKQPREPOVYTLPPSRDEL 537
DB 169 CYSVSSVLPCCAPPMWNGEFTCTAHPLEKTLPLTANITKS-GNTRPEVHLLPPSEEL 227
QY 538 TKKQ-VSLTGLVGFYPSDIAVEMESNGQ--PENNYKTPPVLD-DG--SFLYGLTY 591
DB 228 ALNELVTLTLANGFSPKDVLVNMLQSGQELPREKILTMASRQPSQGTTFVATSLIKY 287
QY 592 DKSRWQGNVSCSVNHEALHNHYTKSLSPG-----IQLDLTC 632
DB 288 AARDWKKGDTFSQVGHGHEALPLAFTQKTIDRLAGKPTHVNVSVMAEVDGTC 339

RESULT 52
HVC3 HETFR STANDARD; PRT; 393 AA.
AC P23086;
DT 01-NOV-1991 (Rel. 20, Last Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE 1g heavy chain C region (Clone 6121) (Fragment).
OS Heterodontus francisci (Horn shark).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Chondrichthyes;
OC Elasmobranchii; Galeomorphii; Heterodontidae; Heterodontiformes;
OC Heterodontidae; Heterodontus.
OX NCBI_TaxID=7792;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Spine;
RX MEDLINE=88328985; Pubmed=1138109;
RA Kokubu F., Hinds K., Litman R., Shambloet M.J., Litman G.W.;
RT "Complete structure and organization of immunoglobulin heavy chain
constant region genes in a phylogenetically primitive vertebrate.";
RL EMBL J. 7:1979-1988(1988).
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/

```

```

CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: X07782, CAA30615.1; -.
CC PIR: S01852; HVRK1.
CC HSSP: P01842; 7FAB.
CC InterPro: IPR007110; IG-1like.
CC InterPro: IPR003597; IG_C1.
CC InterPro: IPR003006; IG_MHC.
CC Pfam: PF00047; Ig_3.
CC SMART: SM00407; IgC1_2.
CC PROSITE: PS00835; IG_LIKE_3.
CC PROSITE: PS00290; IG_MHC_3.
CC Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
FT DOMAIN 1 157 IG-LIKE 1.
FT DOMAIN 168 260 IG-LIKE 2.
FT DOMAIN 270 370 IG-LIKE 3.
FT CARBOHYD 119 119 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 155 155 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 200 200 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 230 230 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 329 329 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 370 370 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 380 380 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 393 AA; 43081 MM; 4E44D076972F18B5 CRC64;

Query Match
Best Local Similarity 27.2%; Pred. No. 2.2e-11;
Matches 118; Conservative 63; Mismatches 162; Indels 91; Gaps 23;

QY 211 VYKKEGEVPEFPFLATFYETKLTGSGELMWQABRASSSWITFDJKNKEVSKRTQDP 270
DB 13 VANKGTYTQ--SSQLTT-ESEVSSKTYCEVRGGS--VW1-----KEI-----PDC 56
QY 271 KLQMGKRLPLHTLTPQALPOVAGSGLTLALBA-----KTGKLHGVNLV 316
DB 57 K--GKVNHPYVILVQSSSEELTSRFAVLVLSIIDFHESITVSLKQGG-HMSEGFPT 112
QY 317 MRATOLQNLTCVEWGPTSPKMLSLKLENKAKSKREKPPVWVNLPEAGMOCCLSDSG 376
DB 113 -----SPTGVNGTFSATSRLT-----VPARE--WFTNK--VYTCQVSHQG 149
QY 377 QVLESNIKVLPTWSTPPVPCPAPPEPKSCDKTHTCPELIGSPVFLPPPKDTLMISRP 436
DB 150 -VTQSEKIT--GSQVPC-----SCN-----DPYKLPESIBOVL-LEATV 186
QY 437 EYTCVVVDVSHEDPEVKFNNYVVDGEVHNAAKTRPEEQYNSTYRVSVLTVLHQMNLNGK 496
DB 187 TLNCT---VSNAFYGVNWM---TQEQKSLKSEIANQPEBDADSVISTVNIISQAWLSGA 240
QY 497 EYKCKVSNKALPAPLEKTIISKAKQD-REPQV-YTLPPSRDELTKQVSLTCLVKCFPYS 554
DB 241 EFYCVVNHODLPTPLASISHKKEVVKDLREPSVSIILSPAEVDVSAQRFSLTCLVRGFSRP 300
QY 555 DIAVEMESNGQPEN--NYKTTTPVLDSD--GSFLYSLKLVVDKSRWQGNVSCVNHGAL 611
DB 301 EIFVKWTINDKSVNPGNYNTEVMENQNSIFYSLSLIAAEWASGASYSCLVGHGAI 360
QY 612 HNNHYTQSLSLSPG 625
DB 361 PLKIRIRTVNKKSSG 374

RESULT 53
ALC1 GORGO STANDARD; PRT; 353 AA.
AC P20758;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE 1g alpha-1 chain C region.
GN IGHAI.

```

OS Gorilla gorilla gorilla (Lowland gorilla).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Gorilla.  
 CC NCBI\_TaxID=9595;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Lymph node;  
 RX MEDLINE=89386006; PubMed=2506527;  
 RA Kawamura S., Omoto K., Ueda S.;  
 RT "Nucleotide sequence of the gorilla immunoglobulin alpha 1 gene";  
 RL Nucleic Acids Res. 17:6732-6732(1989).  
 CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body  
 CC secretions. It may serve both to defend against local infection  
 CC and to prevent access of foreign antigens to the general  
 CC immunologic system.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to: [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X15045; CA33147.1; ALT\_INIT.  
 CC HSSP; P01810; 2PBJ.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003597; Ig\_L1.  
 DR InterPro; IPR003006; Ig\_MHC.  
 DR Pfam; PF00047; Ig; 3.  
 DR SMART; SM00407; IGc1; 2.  
 DR PROSITE; PS00835; IG\_LIKE; 3.  
 DR PROSITE; PS00290; IG\_MHC; 1.  
 KM Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.  
 FT NON\_TER 1 1  
 FT DOMAIN 6 98  
 FT DOMAIN 125 220  
 FT DOMAIN 228 330  
 FT DISULFID 14 14  
 FT DISULFID 26 85  
 FT DISULFID 77 101  
 FT DISULFID 122 122  
 FT DISULFID 123 180  
 FT DISULFID 147 204  
 FT DISULFID 182 182  
 FT DISULFID 192 192  
 FT DISULFID 250 313  
 FT DISULFID 352 352  
 FT CARBOHYD 144 144  
 FT CARBOHYD 340 340  
 SO SEQUENCE 353 AA; 37755 MW; 48208BD802AC514 CRC64;  
 Query Match 8.3%; Score 282.5; DB 1; Length 353;  
 Beest Local Similarity 26.5%; Pred. No. 5.2e-11;  
 Matches 92; Conservative 46; Mismatches 136; Indels 73; Gaps 12;  
 Oy TGKLGHEVNLVVRATOLG--KNLTCEVWSPSPKMLSLKLNKEAVSRKRPVWLN 362  
 Db SGLDTTSSQLTLPPTQCPDGSVCHNHNTP----- 93  
 Oy PEAGWMOCLSDSGVLLSNIKVLPSTVPCAPAPKSCDKHTCPPELLGSPVFLF 422  
 Db -----SQGVTPCRVPSPPRPPSPSTPTSPPCCH-----PRLSLH 130  
 Oy PPKPKDTLMISRTPEVTCVVVDVSHEDPEVKFNKYVDGEVFNAAKTKREPQVNSTYRV 482  
 Db 131 RPALED-LLLGSEANLTCTLTGL--RDAGVFTWTPSSGK--SAVEGPPERDLGCYSVS 186

Oy 483 SVLTIVLQMDWLNKGEYCKVSNKALPAPIEKTISKAGQPREPOVYTLPSRDELTKNQ- 541  
 Db SVLVLGCAEPNHNHGTFTCTAIVPESKTPLTTLTKS--GNMRPVPVHLPPSEELANEL 245  
 Oy 542 VSLTIVLKGFPYPSDIAVWESNNGQ--PENNYKTTTPVID-SDG--SEFLYSKLTVDXSRW 596  
 Db 246 VTLTCLARGFSPKDVLRVWLQSGOELPREKYLTVASRQEPGQTTTFVAVTSILRVAAEDW 305  
 Oy 597 OQGVNFSGVNHEALHNHYTKSLSLSPG-----LQLEDTG 632  
 Db 306 KKGDTFSCWGHLEALPLAFTQKTLDRLAKRPHVNVSVMAEVDGTC 352  
 RESULT 54  
 ALCL\_HUMAN  
 ID ALCL\_HUMAN STANDARD; PRT; 353 AA.  
 AC P01876;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 01-FEB-1991 (Rel. 17, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Ig alpha-1 chain C region.  
 GN IGHAI.  
 OS Homo sapiens (Human).  
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 CC NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=84130179; PubMed=6421489;  
 RA Flanagan J.G., Lefranc M.-P., Rabbits T.H.;  
 RT "Mechanisms of divergence and convergence of the human immunoglobulin  
 RT alpha 1 and alpha 2 constant region gene sequences";  
 RL Cell 36:681-688(1984).  
 RN [2]  
 RP SEQUENCE (MYELOMA PROTEIN BUR), AND DISULFIDE BONDS.  
 RX MEDLINE=79151016; PubMed=107164;  
 RA Putnam F.W., Liu Y.-S.V., Low T.L.K.;  
 RT "Primary structure of a human IgM immunoglobulin. IV. Streptococcal  
 RT IgM protease, digestion, Fab and Fc fragments, and the complete  
 RT amino acid sequence of the alpha 1 heavy chain";  
 RL J. Biol. Chem. 254:2865-2874(1979).  
 RN [3]  
 RP SEQUENCE (MYELOMA PROTEIN TRO).  
 RX MEDLINE=76023781; PubMed=809331;  
 RA Kretzlin H., Altevogt P., Ruban E., Kortz A., Staroscik K.,  
 RA Hilschmann N.;  
 RT "The primary structure of a monoclonal IgA-immunoglobulin (IgA Tro.),  
 RT II. The amino acid sequence of the H-chain, alpha-type, subgroup III;  
 RT structure of the complete IgA-molecule";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 356:1337-1342(1975).  
 RN [4]  
 RP DISULFIDE BONDS.  
 RX MEDLINE=80114124; PubMed=393607;  
 RA Yang C.-Y., Kretzlin H., Gotz H., Hilschmann N.;  
 RT "Rule of antibody structure. Primary structure of a human monoclonal  
 RT IgA-immunoglobulin (myeloma protein Tro). VII. Purification and  
 RT characterization of the disulfide bridges";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 360:1919-1940(1979).  
 RN [5]  
 RP REVIEW.  
 RX MEDLINE=91054387; PubMed=2241915;  
 RA Kerr M.A.;  
 RT "The structure and function of human IgA";  
 RL Biochem. J. 271:285-296(1990).  
 CC -1- FUNCTION: Ig alpha is the major immunoglobulin class in body  
 CC secretions. It may serve both to defend against local infection  
 CC and to prevent access of foreign antigens to the general  
 CC immunologic system.  
 CC -1- SUBUNIT: Monomeric or polymeric.  
 CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -



```

Db 181 -----TOBOKPLSEIAVOPGEDPDSVISTVDISTDAMLSSEAFVGVSHODLPTPLR 233
Oy 513 KTSK-AGGPREPOVYTLPPSRDELTKNQ--VSLTCLVKGYPSDIAEMESNQPEN-- 568
Db 234 DSIHKAADLREPSVSLPPAEISARFLSLTCLVGRGSPREIFVKMTVNDKSVNPG 293
Oy 569 NYKTPPLVDSG-SGFFLYSKLTVDSKRWQGNVFCSCVMHEALHNHTOKSLSPG 625
Db 294 NYKTEVAVENDKSSFFIYLSLSTAEMASGASISCVGHAFILKINRTVNSG 351

RESULT 56
MUCM ICTPU STANDARD; PRT; 481 AA.
AC P23735;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig mu chain C region membrane-bound form.
OS Ictalurus punctatus (Channel catfish).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Siluriformes;
OC Ictaluridae; Ictalurus.
OC NCBI_TaxID=7998;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90384824; PubMed=2119496;
RA Wilson M.R., Marcuz A., van Ginkel F., Miller N.W., Clem L.W.,
RA Middelton D., Warr G.W.;
RT "The immunoglobulin M heavy chain constant region gene of the channel
RT catfish, Ictalurus punctatus: an unusual mRNA splice pattern produces
RT the membrane form of the molecule."
RL Nucleic Acids Res. 18:5227-5233(1990).
CC -I- MISCELLANEOUS: During differentiation, B lymphocytes switch from
CC expression of membrane-bound Igm to secretion of Igm. The mu
CC chain of membrane and secreted Igm differ in their C-terminal
CC segments.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52617; -; NOT_ANNOTATED_CDS.
DR HSBP; P01857; IFC1.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig; 3.
DR SMART; SM00407; IGC1; 1.
DR PROSITE; PS50835; IG_LIKE; 3.
DR PROSITE; PS00290; IG_MHC; 2.
KW Immunoglobulin domain; Immunoglobulin C region; Glycoprotein;
KW Transmembrane.
KW NON_TER
FT 1 101 CH1.
FT 2 102 CH2.
FT 3 103 CH3.
FT 4 104 CH4.
FT 5 105 CH4.
FT 6 106 CH4.
FT 7 107 CH4.
FT 8 108 CH4.
FT 9 109 CH4.
FT 10 110 CH4.
FT 11 111 CH4.
FT 12 112 CH4.
FT 13 113 CH4.
FT 14 114 CH4.
FT 15 115 CH4.
FT 16 116 CH4.
FT 17 117 CH4.
FT 18 118 CH4.
FT 19 119 CH4.
FT 20 120 CH4.
FT 21 121 CH4.
FT 22 122 CH4.
FT 23 123 CH4.
FT 24 124 CH4.
FT 25 125 CH4.
FT 26 126 CH4.
FT 27 127 CH4.
FT 28 128 CH4.
FT 29 129 CH4.
FT 30 130 CH4.
FT 31 131 CH4.
FT 32 132 CH4.
FT 33 133 CH4.
FT 34 134 CH4.
FT 35 135 CH4.
FT 36 136 CH4.
FT 37 137 CH4.
FT 38 138 CH4.
FT 39 139 CH4.
FT 40 140 CH4.
FT 41 141 CH4.
FT 42 142 CH4.
FT 43 143 CH4.
FT 44 144 CH4.
FT 45 145 CH4.
FT 46 146 CH4.
FT 47 147 CH4.
FT 48 148 CH4.
FT 49 149 CH4.
FT 50 150 CH4.
FT 51 151 CH4.
FT 52 152 CH4.
FT 53 153 CH4.
FT 54 154 CH4.
FT 55 155 CH4.
FT 56 156 CH4.
FT 57 157 CH4.
FT 58 158 CH4.
FT 59 159 CH4.
FT 60 160 CH4.
FT 61 161 CH4.
FT 62 162 CH4.
FT 63 163 CH4.
FT 64 164 CH4.
FT 65 165 CH4.
FT 66 166 CH4.
FT 67 167 CH4.
FT 68 168 CH4.
FT 69 169 CH4.
FT 70 170 CH4.
FT 71 171 CH4.
FT 72 172 CH4.
FT 73 173 CH4.
FT 74 174 CH4.
FT 75 175 CH4.
FT 76 176 CH4.
FT 77 177 CH4.
FT 78 178 CH4.
FT 79 179 CH4.
FT 80 180 CH4.
FT 81 181 CH4.
FT 82 182 CH4.
FT 83 183 CH4.
FT 84 184 CH4.
FT 85 185 CH4.
FT 86 186 CH4.
FT 87 187 CH4.
FT 88 188 CH4.
FT 89 189 CH4.
FT 90 190 CH4.
FT 91 191 CH4.
FT 92 192 CH4.
FT 93 193 CH4.
FT 94 194 CH4.
FT 95 195 CH4.
FT 96 196 CH4.
FT 97 197 CH4.
FT 98 198 CH4.
FT 99 199 CH4.
FT 100 200 CH4.
FT 101 201 CH4.
FT 102 202 CH4.
FT 103 203 CH4.
FT 104 204 CH4.
FT 105 205 CH4.
FT 106 206 CH4.
FT 107 207 CH4.
FT 108 208 CH4.
FT 109 209 CH4.
FT 110 210 CH4.
FT 111 211 CH4.
FT 112 212 CH4.
FT 113 213 CH4.
FT 114 214 CH4.
FT 115 215 CH4.
FT 116 216 CH4.
FT 117 217 CH4.
FT 118 218 CH4.
FT 119 219 CH4.
FT 120 220 CH4.
FT 121 221 CH4.
FT 122 222 CH4.
FT 123 223 CH4.
FT 124 224 CH4.
FT 125 225 CH4.
FT 126 226 CH4.
FT 127 227 CH4.
FT 128 228 CH4.
FT 129 229 CH4.
FT 130 230 CH4.
FT 131 231 CH4.
FT 132 232 CH4.
FT 133 233 CH4.
FT 134 234 CH4.
FT 135 235 CH4.
FT 136 236 CH4.
FT 137 237 CH4.
FT 138 238 CH4.
FT 139 239 CH4.
FT 140 240 CH4.
FT 141 241 CH4.
FT 142 242 CH4.
FT 143 243 CH4.
FT 144 244 CH4.
FT 145 245 CH4.
FT 146 246 CH4.
FT 147 247 CH4.
FT 148 248 CH4.
FT 149 249 CH4.
FT 150 250 CH4.
FT 151 251 CH4.
FT 152 252 CH4.
FT 153 253 CH4.
FT 154 254 CH4.
FT 155 255 CH4.
FT 156 256 CH4.
FT 157 257 CH4.
FT 158 258 CH4.
FT 159 259 CH4.
FT 160 260 CH4.
FT 161 261 CH4.
FT 162 262 CH4.
FT 163 263 CH4.
FT 164 264 CH4.
FT 165 265 CH4.
FT 166 266 CH4.
FT 167 267 CH4.
FT 168 268 CH4.
FT 169 269 CH4.
FT 170 270 CH4.
FT 171 271 CH4.
FT 172 272 CH4.
FT 173 273 CH4.
FT 174 274 CH4.
FT 175 275 CH4.
FT 176 276 CH4.
FT 177 277 CH4.
FT 178 278 CH4.
FT 179 279 CH4.
FT 180 280 CH4.
FT 181 281 CH4.
FT 182 282 CH4.
FT 183 283 CH4.
FT 184 284 CH4.
FT 185 285 CH4.
FT 186 286 CH4.
FT 187 287 CH4.
FT 188 288 CH4.
FT 189 289 CH4.
FT 190 290 CH4.
FT 191 291 CH4.
FT 192 292 CH4.
FT 193 293 CH4.
FT 194 294 CH4.
FT 195 295 CH4.
FT 196 296 CH4.
FT 197 297 CH4.
FT 198 298 CH4.
FT 199 299 CH4.
FT 200 300 CH4.
FT 201 301 CH4.
FT 202 302 CH4.
FT 203 303 CH4.
FT 204 304 CH4.
FT 205 305 CH4.
FT 206 306 CH4.
FT 207 307 CH4.
FT 208 308 CH4.
FT 209 309 CH4.
FT 210 310 CH4.
FT 211 311 CH4.
FT 212 312 CH4.
FT 213 313 CH4.
FT 214 314 CH4.
FT 215 315 CH4.
FT 216 316 CH4.
FT 217 317 CH4.
FT 218 318 CH4.
FT 219 319 CH4.
FT 220 320 CH4.
FT 221 321 CH4.
FT 222 322 CH4.
FT 223 323 CH4.
FT 224 324 CH4.
FT 225 325 CH4.
FT 226 326 CH4.
FT 227 327 CH4.
FT 228 328 CH4.
FT 229 329 CH4.
FT 230 330 CH4.
FT 231 331 CH4.
FT 232 332 CH4.
FT 233 333 CH4.
FT 234 334 CH4.
FT 235 335 CH4.
FT 236 336 CH4.
FT 237 337 CH4.
FT 238 338 CH4.
FT 239 339 CH4.
FT 240 340 CH4.
FT 241 341 CH4.
FT 242 342 CH4.
FT 243 343 CH4.
FT 244 344 CH4.
FT 245 345 CH4.
FT 246 346 CH4.
FT 247 347 CH4.
FT 248 348 CH4.
FT 249 349 CH4.
FT 250 350 CH4.
FT 251 351 CH4.
FT 252 352 CH4.
FT 253 353 CH4.
FT 254 354 CH4.
FT 255 355 CH4.
FT 256 356 CH4.
FT 257 357 CH4.
FT 258 358 CH4.
FT 259 359 CH4.
FT 260 360 CH4.
FT 261 361 CH4.
FT 262 362 CH4.
FT 263 363 CH4.
FT 264 364 CH4.
FT 265 365 CH4.
FT 266 366 CH4.
FT 267 367 CH4.
FT 268 368 CH4.
FT 269 369 CH4.
FT 270 370 CH4.
FT 271 371 CH4.
FT 272 372 CH4.
FT 273 373 CH4.
FT 274 374 CH4.
FT 275 375 CH4.
FT 276 376 CH4.
FT 277 377 CH4.
FT 278 378 CH4.
FT 279 379 CH4.
FT 280 380 CH4.
FT 281 381 CH4.
FT 282 382 CH4.
FT 283 383 CH4.
FT 284 384 CH4.
FT 285 385 CH4.
FT 286 386 CH4.
FT 287 387 CH4.
FT 288 388 CH4.
FT 289 389 CH4.
FT 290 390 CH4.
FT 291 391 CH4.
FT 292 392 CH4.
FT 293 393 CH4.
FT 294 394 CH4.
FT 295 395 CH4.
FT 296 396 CH4.
FT 297 397 CH4.
FT 298 398 CH4.
FT 299 399 CH4.
FT 300 400 CH4.
FT 301 401 CH4.
FT 302 402 CH4.
FT 303 403 CH4.
FT 304 404 CH4.
FT 305 405 CH4.
FT 306 406 CH4.
FT 307 407 CH4.
FT 308 408 CH4.
FT 309 409 CH4.
FT 310 410 CH4.
FT 311 411 CH4.
FT 312 412 CH4.
FT 313 413 CH4.
FT 314 414 CH4.
FT 315 415 CH4.
FT 316 416 CH4.
FT 317 417 CH4.
FT 318 418 CH4.
FT 319 419 CH4.
FT 320 420 CH4.
FT 321 421 CH4.
FT 322 422 CH4.
FT 323 423 CH4.
FT 324 424 CH4.
FT 325 425 CH4.
FT 326 426 CH4.
FT 327 427 CH4.
FT 328 428 CH4.
FT 329 429 CH4.
FT 330 430 CH4.
FT 331 431 CH4.
FT 332 432 CH4.
FT 333 433 CH4.
FT 334 434 CH4.
FT 335 435 CH4.
FT 336 436 CH4.
FT 337 437 CH4.
FT 338 438 CH4.
FT 339 439 CH4.
FT 340 440 CH4.
FT 341 441 CH4.
FT 342 442 CH4.
FT 343 443 CH4.
FT 344 444 CH4.
FT 345 445 CH4.
FT 346 446 CH4.
FT 347 447 CH4.
FT 348 448 CH4.
FT 349 449 CH4.
FT 350 450 CH4.
FT 351 451 CH4.
FT 352 452 CH4.
FT 353 453 CH4.
FT 354 454 CH4.
FT 355 455 CH4.
FT 356 456 CH4.
FT 357 457 CH4.
FT 358 458 CH4.
FT 359 459 CH4.
FT 360 460 CH4.
FT 361 461 CH4.
FT 362 462 CH4.
FT 363 463 CH4.
FT 364 464 CH4.
FT 365 465 CH4.
FT 366 466 CH4.
FT 367 467 CH4.
FT 368 468 CH4.
FT 369 469 CH4.
FT 370 470 CH4.
FT 371 471 CH4.
FT 372 472 CH4.
FT 373 473 CH4.
FT 374 474 CH4.
FT 375 475 CH4.
FT 376 476 CH4.
FT 377 477 CH4.
FT 378 478 CH4.
FT 379 479 CH4.
FT 380 480 CH4.
FT 381 481 CH4.
FT 382 482 CH4.
FT 383 483 CH4.
FT 384 484 CH4.
FT 385 485 CH4.
FT 386 486 CH4.
FT 387 487 CH4.
FT 388 488 CH4.
FT 389 489 CH4.
FT 390 490 CH4.
FT 391 491 CH4.
FT 392 492 CH4.
FT 393 493 CH4.
FT 394 494 CH4.
FT 395 495 CH4.
FT 396 496 CH4.
FT 397 497 CH4.
FT 398 498 CH4.
FT 399 499 CH4.
FT 400 500 CH4.
FT 401 501 CH4.
FT 402 502 CH4.
FT 403 503 CH4.
FT 404 504 CH4.
FT 405 505 CH4.
FT 406 506 CH4.
FT 407 507 CH4.
FT 408 508 CH4.
FT 409 509 CH4.
FT 410 510 CH4.
FT 411 511 CH4.
FT 412 512 CH4.
FT 413 513 CH4.
FT 414 514 CH4.
FT 415 515 CH4.
FT 416 516 CH4.
FT 417 517 CH4.
FT 418 518 CH4.
FT 419 519 CH4.
FT 420 520 CH4.
FT 421 521 CH4.
FT 422 522 CH4.
FT 423 523 CH4.
FT 424 524 CH4.
FT 425 525 CH4.
FT 426 526 CH4.
FT 427 527 CH4.
FT 428 528 CH4.
FT 429 529 CH4.
FT 430 530 CH4.
FT 431 531 CH4.
FT 432 532 CH4.
FT 433 533 CH4.
FT 434 534 CH4.
FT 435 535 CH4.
FT 436 536 CH4.
FT 437 537 CH4.
FT 438 538 CH4.
FT 439 539 CH4.
FT 440 540 CH4.
FT 441 541 CH4.
FT 442 542 CH4.
FT 443 543 CH4.
FT 444 544 CH4.
FT 445 545 CH4.
FT 446 546 CH4.
FT 447 547 CH4.
FT 448 548 CH4.
FT 449 549 CH4.
FT 450 550 CH4.
FT 451 551 CH4.
FT 452 552 CH4.
FT 453 553 CH4.
FT 454 554 CH4.
FT 455 555 CH4.
FT 456 556 CH4.
FT 457 557 CH4.
FT 458 558 CH4.
FT 459 559 CH4.
FT 460 560 CH4.
FT 461 561 CH4.
FT 462 562 CH4.
FT 463 563 CH4.
FT 464 564 CH4.
FT 465 565 CH4.
FT 466 566 CH4.
FT 467 567 CH4.
FT 468 568 CH4.
FT 469 569 CH4.
FT 470 570 CH4.
FT 471 571 CH4.
FT 472 572 CH4.
FT 473 573 CH4.
FT 474 574 CH4.
FT 475 575 CH4.
FT 476 576 CH4.
FT 477 577 CH4.
FT 478 578 CH4.
FT 479 579 CH4.
FT 480 580 CH4.
FT 481 581 CH4.
FT 482 582 CH4.
FT 483 583 CH4.
FT 484 584 CH4.
FT 485 585 CH4.
FT 486 586 CH4.
FT 487 587 CH4.
FT 488 588 CH4.
FT 489 589 CH4.
FT 490 590 CH4.
FT 491 591 CH4.
FT 492 592 CH4.
FT 493 593 CH4.
FT 494 594 CH4.
FT 495 595 CH4.
FT 496 596 CH4.
FT 497 597 CH4.
FT 498 598 CH4.
FT 499 599 CH4.
FT 500 600 CH4.
FT 501 601 CH4.
FT 502 602 CH4.
FT 503 603 CH4.
FT 504 604 CH4.
FT 505 605 CH4.
FT 506 606 CH4.
FT 507 607 CH4.
FT 508 608 CH4.
FT 509 609 CH4.
FT 510 610 CH4.
FT 511 611 CH4.
FT 512 612 CH4.
FT 513 613 CH4.
FT 514 614 CH4.
FT 515 615 CH4.
FT 516 616 CH4.
FT 517 617 CH4.
FT 518 618 CH4.
FT 519 619 CH4.
FT 520 620 CH4.
FT 521 621 CH4.
FT 522 622 CH4.
FT 523 623 CH4.
FT 524 624 CH4.
FT 525 625 CH4.
FT 526 626 CH4.
FT 527 627 CH4.
FT 528 628 CH4.
FT 529 629 CH4.
FT 530 630 CH4.
FT 531 631 CH4.
FT 532 632 CH4.
FT 533 633 CH4.
FT 534 634 CH4.
FT 535 635 CH4.
FT 536 636 CH4.
FT 537 637 CH4.
FT 538 638 CH4.
FT 539 639 CH4.
FT 540 640 CH4.
FT 541 641 CH4.
FT 542 642 CH4.
FT 543 643 CH4.
FT 544 644 CH4.
FT 545 645 CH4.
FT 546 646 CH4.
FT 547 647 CH4.
FT 548 648 CH4.
FT 549 649 CH4.
FT 550 650 CH4.
FT 551 651 CH4.
FT 552 652 CH4.
FT 553 653 CH4.
FT 554 654 CH4.
FT 555 655 CH4.
FT 556 656 CH4.
FT 557 657 CH4.
FT 558 658 CH4.
FT 559 659 CH4.
FT 560 660 CH4.
FT 561 661 CH4.
FT 562 662 CH4.
FT 563 663 CH4.
FT 564 664 CH4.
FT 565 665 CH4.
FT 566 666 CH4.
FT 567 667 CH4.
FT 568 668 CH4.
FT 569 669 CH4.
FT 570 670 CH4.
FT 571 671 CH4.
FT 572 672 CH4.
FT 573 673 CH4.
FT 574 674 CH4.
FT 575 675 CH4.
FT 576 676 CH4.
FT 577 677 CH4.
FT 578 678 CH4.
FT 579 679 CH4.
FT 580 680 CH4.
FT 581 681 CH4.
FT 582 682 CH4.
FT 583 683 CH4.
FT 584 684 CH4.
FT 585 685 CH4.
FT 586 686 CH4.
FT 587 687 CH4.
FT 588 688 CH4.
FT 589 689 CH4.
FT 590 690 CH4.
FT 591 691 CH4.
FT 592 692 CH4.
FT 593 693 CH4.
FT 594 694 CH4.
FT 595 695 CH4.
FT 596 696 CH4.
FT 597 697 CH4.
FT 598 698 CH4.
FT 599 699 CH4.
FT 600 700 CH4.
FT 601 701 CH4.
FT 602 702 CH4.
FT 603 703 CH4.
FT 604 704 CH4.
FT 605 705 CH4.
FT 606 706 CH4.
FT 607 707 CH4.
FT 608 708 CH4.
FT 609 709 CH4.
FT 610 710 CH4.
FT 611 711 CH4.
FT 612 712 CH4.
FT 613 713 CH4.
FT 614 714 CH4.
FT 615 715 CH4.
FT 616 716 CH4.
FT 617 717 CH4.
FT 618 718 CH4.
FT 619 719 CH4.
FT 620 720 CH4.
FT 621 721 CH4.
FT 622 722 CH4.
FT 623 723 CH4.
FT 624 724 CH4.
FT 625 725 CH4.
FT 626 726 CH4.
FT 627 727 CH4.
FT 628 728 CH4.
FT 629 729 CH4.
FT 630 730 CH4.
FT 631 731 CH4.
FT 632 732 CH4.
FT 633 733 CH4.
FT 634 734 CH4.
FT 635 735 CH4.
FT 636 736 CH4.
FT 637 737 CH4.
FT 638 738 CH4.
FT 639 739 CH4.
FT 640 740 CH4.
FT 641 741 CH4.
FT 642 742 CH4.
FT 643 743 CH4.
FT 644 744 CH4.
FT 645 745 CH4.
FT 646 746 CH4.
FT 647 747 CH4.
FT 648 748 CH4.
FT 649 749 CH4.
FT 650 750 CH4.
FT 651 751 CH4.
FT 652 752 CH4.
FT 653 753 CH4.
FT 654 754 CH4.
FT 655 755 CH4.
FT 656 756 CH4.
FT 657 757 CH4.
FT 658 758 CH4.
FT 659 759 CH4.
FT 660 760 CH4.
FT 661 761 CH4.
FT 662 762 CH4.
FT 663 763 CH4.
FT 664 764 CH4.
FT 665 765 CH4.
FT 666 766 CH4.
FT 667 767 CH4.
FT 668 768 CH4.
FT 669 769 CH4.
FT 670 770 CH4.
FT 671 771 CH4.
FT 672 772 CH4.
FT 673 773 CH4.
FT 674 774 CH4.
FT 675 775 CH4.
FT 676 776 CH4.
FT 677 777 CH4.
FT 678 778 CH4.
FT 679 779 CH4.
FT 680 780 CH4.
FT 681 781 CH4.
FT 682 782 CH4.
FT 683 783 CH4.
FT 684 784 CH4.
FT 685 785 CH4.
FT 686 786 CH4.
FT 687 787 CH4.
FT 688 788 CH4.
FT 689 789 CH4.
FT 690 790 CH4.
FT 691 791 CH4.
FT 692 792 CH4.
FT 693 793 CH4.
FT 694 794 CH4.
FT 695 795 CH4.
FT 696 796 CH4.
FT 697 797 CH4.
FT 698 798 CH4.
FT 699 799 CH4.
FT 700 800 CH4.
FT 701 801 CH4.
FT 702 802 CH4.
FT 703 803 CH4.
FT 704 804 CH4.
FT 705 805 CH4.
FT 706 806 CH4.
FT 707 807 CH4.
FT 708 808 CH4.
FT 709 809 CH4.
FT 710 810 CH4.
FT 711 811 CH4.
FT 712 812 CH4.
FT 713 813 CH4.
FT 714 814 CH4.
FT 715 815 CH4.
FT 716 816 CH4.
FT 717 817 CH4.
FT 718 818 CH4.
FT 719 819 CH4.
FT 720 820 CH4.
FT 721 821 CH4.
FT 722 822 CH4.
FT 723 823 CH4.
FT 724 824 CH4.
FT 725 825 CH4.
FT 726 826 CH4.
FT 727 827 CH4.
FT 728 828 CH4.
FT 729 829 CH4.
FT 730 830 CH4.
FT 731 831 CH4.
FT 732 832 CH4.
FT 733 833 CH4.
FT 734 834 CH4.
FT 735 835 CH4.
FT 736 836 CH4.
FT 737 837 CH4.
FT 738 838 CH4.
FT 739 839 CH4.
FT 740 840 CH4.
FT 741 841 CH4.
FT 742 842 CH4.
FT 743 843 CH4.
FT 744 844 CH4.
FT 745 845 CH4.
FT 746 846 CH4.
FT 747 847 CH4.
FT 748 848 CH4.
FT 749 849 CH4.
FT 750 850 CH4.
FT 751 851 CH4.
FT 752 852 CH4.
FT 753 853 CH4.
FT 754 854 CH4.
FT 755 855 CH4.
FT 756 856 CH4.
FT 757 857 CH4.
FT 758 858 CH4.
FT 759 859 CH4.
FT 760 860 CH4.
FT 761 861 CH4.
FT 762 862 CH4.
FT 763 863 CH4.
FT 764 864 CH4.
FT 765 865 CH4.
FT 766 866 CH4.
FT 767 867 CH4.
FT 768 868 CH4.
FT 769 869 CH4.
FT 770 870 CH4.
FT 771 871 CH4.
FT 772 872 CH4.
FT 773 873 CH4.
FT 774 874 CH4.
FT 775 875 CH4.
FT 776 876 CH4.
FT 777 877 CH4.
FT 778 878 CH4.
FT 779 879 CH4.
FT 780 880 CH4.
FT 781 881 CH4.
FT 782 882 CH4.
FT 783 883 CH4.
FT 784 884 CH4.
FT 785 885 CH4.
FT 786 886 CH4.
FT 787 887 CH4.
FT 788 888 CH4.
FT 789 889 CH4.
FT 790 890 CH4.
FT 791 891 CH4.
FT 792 892 CH4.
FT 793 893 CH4.
FT 794 894 CH4.
FT 795 895 CH4.
FT 796 896 CH4.
FT 797 897 CH4.
FT 798 898 CH4.
FT 799 899 CH4.
FT 800 900 CH4.
FT 801 901 CH4.
FT 802 902 CH4.
FT 803 903 CH4.
FT 804 904 CH4.
FT 805 905 CH4.
FT 806 906 CH4.
FT 807 907 CH4.
FT 808 908 CH4.
FT 809 909 CH4.
FT 810 910 CH4.
FT 811 911 CH4.
FT 812 912 CH4.
FT 813 913 CH4.
FT 814 914 CH4.
FT 815 915 CH4.
FT 816 916 CH4.
FT 817 917 CH4.
FT 818 918 CH4.
FT 819 919 CH4.
FT 820 920 CH4.
FT 821 921 CH4.
FT 822 922 CH4.
FT 823 923 CH4.
FT 824 924 CH4.
FT 825 925 CH4.
FT 826 926 CH4.
FT 827 927 CH4.
FT 828 928 CH4.
FT 829 929 CH4.
FT 830 930 CH4.
FT 831 931 CH4.
FT 832 932 CH4.
FT 833 933 CH4.
FT 834 934 CH4.
FT 835 935 CH4.
FT 836 936 CH4.
FT 837 937 CH4.
FT 838 938 CH4.
FT 839 939 CH4.
FT 840 940 CH4.
FT 841 941 CH4.
FT 842 942 CH4.
FT 843 943 CH4.
FT 844 944 CH4.
FT 845 945 CH4.
FT 846 946 CH4.
FT 847 947 CH4.
FT 848 948 CH4.
FT 849 949 CH4.
FT 850 950 CH4.
FT 851 951 CH4.
FT 852 952 CH4.
FT 853 953 CH4.
FT 854 954 CH4.
FT 855 955 CH4.
FT 856 956 CH4.
FT 857 957 CH4.
FT 858 958 CH4.
FT 859 959 CH4.
FT 860 960 CH4.
FT 861 961 CH4.
FT 862 962 CH4.
FT 863 963 CH4.
FT 864 964 CH4.
FT 865 965 CH4.
FT 866 966 CH4.
FT 867 967 CH4.
FT 868 968 CH4.
FT 869 969 CH4.
FT 870 970 CH4.
FT 871 971 CH4.
FT 872 972 CH4.
FT 873 973 CH4.
FT 874 974 CH4.
FT 875 975 CH4.
FT 876 976 CH4.
FT 877 977 CH4.
FT 878 978 CH4.
FT 879 979 CH
```

```
CC secretions. It may serve both to defend against local infection
CC and to prevent access of foreign antigens to the general
CC immunologic system.
CC -1- MISCELLANEOUS: The final C-region domain is deleted from Ref. 2
CC chain. It was isolated from a myeloma protein that contains 1
CC light and 1 heavy chain per molecule, linked by a disulfide bond.
CC In contrast, normal mouse IgA molecules contain 2 light and 2
CC heavy chains and lack a light-heavy chain disulfide bond.
CC -1- MISCELLANEOUS: M511 chain was isolated from a myeloma protein that
CC binds phosphorylcholine.
CC -1- MISCELLANEOUS: M511 sequence was compared with that of mouse MOPC
CC 47A, and a genetic mechanism for the deletion of the CH3 domain of
CC the mutant chain is proposed.
CC -1- SIMILARITY: Contains 3 immunoglobulin-like domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: D11468; BAA02026.1; ALT_INIT.
CC PIR: A91479; AHMS.
CC HSSP: P01810; 2FBU.
CC InterPro: IPR007110; Ig-like.
CC InterPro: IPR003597; Ig_c1.
CC InterPro: IPR003006; Ig_MHC.
CC Pfam: PF00047; Ig; 3.
CC SMART: SM00407; Igcl; 2.
CC PROSITE: PSS0835; IG_LIKE; 3.
CC PROSITE: PS00290; IG_MHC; 2.
CC Immunoglobulin domain; Immunoglobulin C region; Glycoprotein; Repeat.
CC KW NON_TER 1
CC FT DOMAIN 1 99 IG-LIKE 1.
CC FT DOMAIN 6 206 IG-LIKE 2.
CC FT DOMAIN 219 321 IG-LIKE 3.
CC FT DISULFID 26 84 BY SIMILARITY.
CC FT DISULFID 76 100 BY SIMILARITY.
CC FT DISULFID 114 171 BY SIMILARITY.
CC FT DISULFID 138 195 BY SIMILARITY.
CC FT CARBOHYD 38 38 N-LINKED (GLCNAC. . .).
CC FT CARBOHYD 101 101 O-LINKED.
CC FT CARBOHYD 331 331 O-LINKED.
CC FT VARIANT 2 2 S -> A (IN MOPC 47A).
CC FT VARIANT 18 18 S -> C (IN MOPC 47A AND M511).
CC FT VARIANT 67 67 N -> S (IN MOPC 47A AND M511).
CC FT VARIANT 73 73 A -> T (IN MOPC 47A).
CC FT VARIANT 112 112 P -> G (IN M511).
CC FT VARIANT 135 135 S -> Q (IN MOPC 47A AND M511).
CC FT VARIANT 141 141 N -> D (IN MOPC 47A AND M511).
CC FT VARIANT 168 168 Q -> E (IN MOPC 47A).
CC FT VARIANT 212 213 VT -> SQ (IN MOPC 47A).
CC FT VARIANT 235 235 E -> G (IN M511).
CC FT VARIANT 255 255 MISSING (IN M511).
CC FT VARIANT 295 295 T -> D (IN M511).
CC FT VARIANT 301 301 Q -> G (IN M511).
CC FT VARIANT 329 329 N -> Q (IN M511).
CC FT VARIANT 331 331 S -> N (IN M511).
CC SQ SEQUENCE 344 AA; 36876 MW; 3694CFEP9B19A9P8 CRC64;
Query Match 7.4%; Score 251.5; DB 1; Length 344;
Best Local Similarity 25.5%; Pred. NO. 4.8e-09;
Matches 93; Conservative 55; Mismatches 147; Indels 69; Gaps 13;
```

```
QY 391 SMPVCPAPRPKSCDKHTTCPELLGSPVFLFPPKXDTIMSTRTEPVTVVVDVSHEDP 450
DB 101 SPTPTPTPTPTPTSCQ-----PSLSLQRPALPD-LTIGSDASITCTINGLNRPG 148
QY 451 EVKFMVYDGVGVNNAKTPREEDQYNT--YRVSVLTVLHQMVLNGEKYCKVSNKAL 507
DB 149 AV-FTW-----EPSTGKAVQKKAQVNSGCCSVSVLPGCERNNSGASFCTVTHPE- 201
QY 508 PAPIKTKSKAKQGPREFQVYTLPPSRDELTKNQ-VSLTCLVKGFPSPDIAYEWSNQ- 565
DB 202 SGLTGLTIKATVNTVTFPPQVHLLPPPSERIALNLELSTLCVRAVNPKEVLVRWLGND 261
QY 566 --PENNYKTPPVVLDSDG--SPEFLSKLTVDKSRMOGVFSCVWHEALNHYQKSL 621
DB 262 LSPSEVLVFEPLKEPGEKATYLVTSVRVSAETWKQGOVSCMGHEALPNVFTOKID 321
QY 622 LSPG 625
DB 322 RLSG 325
RESULT 58
SHS1 MOUSE STANDARD; PRT; 513 AA.
AC P97757; O08907; O35924; O08555; O85556; P97796; Q08559; Q9QX57;
Q9WTN4;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Protein-tyrosine phosphatase non-receptor type subunit 1 precursor
DE (SHP substrate-1) (SHPs-1) (Inhibitory receptor SHPS-1) (Signal-
DE regulatory protein alpha-1) (SIRP-alpha-1) (MSIRP-alpha) (MyD-1
DE antigen) (Brain Ig-like molecule with tyrosine-based activation
DE motifs) (Bit) (p84).
GN PTNS1 OR SHPS1 OR SIRP OR MYD1 OR BIT.
OC Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_Taxid=10090;
RN (1)
RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC TISSUE=Brain;
RX MEDLINE=97223399; PubMed=9070220;
RA Yamao T., Matczak T., Amano K., Matsuda Y., Takahashi N., Ochi F.,
RA Fujioka Y., Kasuga M.;
RT "Mouse and human SHPS-1: molecular cloning of cDNAs and chromosomal
RT localization of genes.";
RN Biochem. Biophys. Res. Commun. 231:61-67(1997).
RN (2)
RP SEQUENCE FROM N.A. (ISOFORM 2) AND VARIANTS ALA-29; ARG-67; ARG-91;
RP THR-96; SER-128; PRO-194 AND ASN-224.
RC STRAIN=BALB/c; TISSUE=Brain;
RX MEDLINE=97230468; PubMed=9073522;
RA Ohnishi H., Kubota M., Sano S.-I.;
RA "BIT (Bit) maps to mouse chromosome 2.";
RL Genomics 40:504-506(1997).
RN (3)
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), SEQUENCE OF 32-53 AND 422-433
RP (ISOFORM 2), VARIANTS ALA-29; ARG-91; THR-96 AND SER-128,
RP N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
RC STRAIN=BALB/c; TISSUE=Brain, and Cerebellum;
RX MEDLINE=98012243; PubMed=9348339;
RA Comu S., Weng W., Olinick S., Ishwad P., Mi Z., Hempel J., Watkins S.,
RA Lagenaar C.F., Narayanan V.;
RA "The murine p84 neural adhesion molecule is SHPS-1, a member of the
RT phosphatase-binding protein family.";
RL J. Neurosci. 17:8702-8710(1997).
RN (4)
RP SEQUENCE FROM N.A. (ISOFORMS 2 AND 3), VARIANTS ALA-29; ARG-67;
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365,
RP N-GLYCOSYLATION, PHOSPHORYLATION ON TYROSINE RESIDUES, INTERACTION
```

RP WITH PTPN6, AND TISSUE SPECIFICITY.  
RC STRAIN=C57BL/6; TISSUE=fetal thymus;  
RX MEDLINE=96380500; PubMed=9712903;  
RA Veilleux A., Thibaudan E., Latour S.;  
RT "High expression of inhibitory receptor SHP-1 and its association  
RT with protein tyrosine phosphatase SHP-1 in macrophages.";  
RL J. Biol. Chem. 273:22719-22728(1998).  
[5]  
RP SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3), VARIANTS ALA-29; ARG-67;  
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;  
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365, AND  
RP N-GLYCOSYLATION.  
RC STRAIN=129/SvJ, and C57BL/6; TISSUE=Brain, and Liver;  
RX MEDLINE=20053880; PubMed=10585853;  
RA Sano S.-I., Ohnishi H., Kubota M.;  
RT "Gene structure of mouse BTK/SHPs-1.";  
RL Biochem. J. 344:667-675(1999).  
[6]  
RP SEQUENCE FROM N.A. (ISOFORM 2), AND VARIANTS ALA-29; ARG-67;  
RP PRO-74; ALA-83; 86-TYR-VAL-87; ILE-90; ARG-91; THR-96; ALA-114;  
RP ILE-118; SER-128; PRO-194; ASN-224; PHE-351 AND ASP-365.  
RC STRAIN=129/SvJ, and ISS;  
RX MEDLINE=21363810; PubMed=11471062;  
RA Bhinger M.A., Thompson J., Conroy O., Xu Y., Yang F., Canniff J.,  
RA Beeson M., Gordon L., Bennett B., Johnson T.E., Sikele J.M.;  
RT "High-throughput sequence identification of gene coding variants  
RT within alcohol-related OTUs";  
RL Mamm. Genome 12:657-663(2001).  
[7]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANT PRO-194.  
RC TISSUE=Placenta;  
RA Wang H., Chen Z., Ullrich A.;  
RT "Epidermal growth factor-induced association of SHP2 with mouse SIRP-  
RT alpha.";  
RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.  
[8]  
RP FUNCTION, TISSUE SPECIFICITY, AND DEVELOPMENTAL STAGE.  
RX MEDLINE=90152134; PubMed=2203162;  
RA Chuang W., Laganaur C.F.;  
RT "Central nervous system antigen P84 can serve as a substrate for  
RT neurite outgrowth.";  
RL Dev. Biol. 137:219-232(1990).  
[9]  
RP N-GLYCOSYLATION, PHOSPHORYLATION BY JAK2 IN RESPONSE TO GROWTH  
RP HORMONE, AND INTERACTIONS WITH JAK2 AND PTPN11.  
RX MEDLINE=96175985; PubMed=9507023;  
RA Stofega M.R., Wang H., Ullrich A., Carter-Su C.;  
RT "Growth hormone regulation of SIRP and SHP-2 tyrosyl phosphorylation  
RT and association.";  
RL J. Biol. Chem. 273:7112-7117(1998).  
[10]  
RP INTERACTION WITH CD47, AND TISSUE SPECIFICITY.  
RX MEDLINE=99091586; PubMed=9872987;  
RA Jiang P., Laganaur C.F., Narayanan V.;  
RT "Integrin-associated protein is a ligand for the P84 neural adhesion  
RT molecule.";  
RL J. Biol. Chem. 274:559-562(1999).  
[11]  
RP INTERACTION WITH FGR.  
RX MEDLINE=20130295; PubMed=10662797;  
RA Gresham H.D., Dale B.M., Potter J.W., Chang P.W., Vines C.M.,  
RA Lowell C.A., Laganaur C.F., Willman C.L.;  
RT "Negative regulation of phagocytosis in murine macrophages by the Src  
RT kinase family member, Fgr.";  
RL J. Exp. Med. 191:515-528(2000).  
-1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts  
as docking protein and induces translocation of PTPN6.  
PPN1 and other binding partners from the cytosol to the  
plasma membrane. Supports adhesion of cerebellar neurons, neurite  
outgrowth and glial cell attachment. May play a key role in  
intracellular signaling during synaptogenesis and in synaptic  
function. Involved in the negative regulation of receptor tyrosine  
kinase-coupled cellular responses induced by cell adhesion, growth

CC factors or insulin. Mediates negative regulation of phagocytosis,  
CC mast cell activation and dendritic cell activation. CD47 binding  
CC prevents maturation of immature dendritic cells and inhibits  
CC cytokine production by mature dendritic cells (By similarity).  
CC -1- SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in  
CC macrophages, where it primarily binds PTPN6. Binds GRB2 vitro.  
CC Binds FGR. Binds JAK2 irrespective of its phosphorylation status  
CC and forms a stable complex. Binds SCAP1 and/or SCAP2. The  
CC resulting complex recruits Fyb. Binds PTK2B (By similarity).  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC Event=Alternative splicing; Named isoforms=3;  
CC Comment=Additional isoforms seem to exist;  
CC Name=1; Synonyms=a;  
CC IsoId=P97797-1; Sequence=Displayed;  
CC Name=2; Synonyms=a', Large;  
CC IsoId=P97797-2; Sequence=VSP\_007032;  
CC Name=3; Synonyms=b, Small;  
CC IsoId=P97797-3; Sequence=VSP\_007031;  
CC -1- TISSUE SPECIFICITY: Highly expressed in cerebral cortex, brain,  
CC spinal cord, cerebellum and spleen, and at much lower levels in  
CC kidney, thymus, heart, lung and liver. Within the cerebellum,  
CC highly expressed throughout the molecular layer, and in synaptic  
CC glomeruli in the granule cell layer. Detected in neurons of the  
CC hippocampus and dentate gyrus, and in olfactory bulb. Not detected  
CC in Purkinje cells. Highly expressed in the plexiform layers, optic  
CC fiber layer and the outer segments of the photoreceptor layer in  
CC the retina. Highly expressed in macrophages. Isoform 3 is detected  
CC at very low levels in all tissues tested.  
CC -1- DEVELOPMENTAL STAGE: Highly expressed in the CNS of embryos from  
CC day 7 to 17.  
CC -1- PTM: N-glycosylated.  
CC -1- PTM: Phosphorylated on tyrosine residues.  
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.  
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announcements>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL; D87967; BA13520.1; -;  
DR EMBL; D87968; BA13521.1; -;  
DR EMBL; D85785; BA120376.1; -;  
DR EMBL; U89694; AAB92591.1; -;  
DR EMBL; AF072543; AAC24886.1; -;  
DR EMBL; AF072544; AAC24887.1; -;  
DR EMBL; AB024507; BAA89290.1; -;  
DR EMBL; AB024507; BAA89290.1; JOINED.  
DR EMBL; AB024507; BAA89290.1; JOINED.  
DR EMBL; AB024501; BAA89290.1; JOINED.  
DR EMBL; AB024502; BAA89290.1; JOINED.  
DR EMBL; AB024502; BAA89290.1; JOINED.  
DR EMBL; AB024504; BAA89290.1; JOINED.  
DR EMBL; AB024504; BAA89290.1; JOINED.  
DR EMBL; AB024505; BAA89290.1; JOINED.  
DR EMBL; AB024505; BAA89290.1; JOINED.  
DR EMBL; AB024506; BAA89290.1; JOINED.  
DR EMBL; AB024507; BAA89289.1; -;  
DR EMBL; AB024507; BAA89289.1; -;  
DR EMBL; AB024507; BAA89289.1; JOINED.  
DR EMBL; AB024501; BAA89289.1; JOINED.  
DR EMBL; AB024502; BAA89289.1; JOINED.  
DR EMBL; AB024503; BAA89289.1; JOINED.  
DR EMBL; AB024504; BAA89289.1; JOINED.  
DR EMBL; AB024505; BAA89289.1; JOINED.  
DR EMBL; AB024506; BAA89289.1; JOINED.  
DR EMBL; AF332079; AAK56108.1; -;  
DR EMBL; AF332080; AAK56108.1; -;  
DR EMBL; Y10349; CAAT73375.1; -;  
DR PIR; JC5289; JC5289.  
DR MGD; MGI:108563; Pcpn1.  
CO; GO:0008580; F:cytoskeletal regulator activity; IMP.





CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; X62515; CAA54373.1; -

DR EMBL; M85289; AA52700.1; -

DR EMBL; AL445795; CAC18534.1; -

DR EMBL; M64283; AA52699.1; -

DR EMBL; S76436; AA82121.2; -

DR EMBL; L22078; -; NOT\_ANNOTATED\_CDS.

DR PIR; A38096; A38096.

DR HSP; P00740; IEDM.

DR Slens-2DPAGE; P96160; -

DR Genew; HGNC:5273; HSPG2.

DR MIM; 142461; -

DR MIM; 255800; -

DR InterPro; IPR009895; Cona\_like\_1ec\_g1.

DR InterPro; IPR000742; EGF\_2.

DR InterPro; IPR006209; EGF\_like.

DR InterPro; IPR006210; IEGF.

DR InterPro; IPR007110; IG-like.

DR InterPro; IPR003599; IG\_c2.

DR InterPro; IPR003598; IG\_c2.

DR InterPro; IPR003596; IG\_v.

DR InterPro; IPR000034; Laminin\_B.

DR InterPro; IPR002049; Laminin\_EGF.

DR InterPro; IPR001791; Laminin\_G.

DR InterPro; IPR002172; LDL\_receptor\_A.

DR InterPro; IPR000082; SEA\_domain.

DR Pfam; PF00008; EGF; 4.

DR Pfam; PF00047; IG; 22.

DR Pfam; PF00052; Laminin\_B; 3.

DR Pfam; PF00053; Laminin\_EGF; 7.

DR Pfam; PF00054; Laminin\_G; 3.

DR Pfam; PF00057; Idl\_recep; 4.

DR Pfam; PF01390; SEA; 1.

DR PRINTS; PR00261; LDLRECEPTOR.

DR PRODOM; PD003031; Laminin\_B; 3.

DR SMART; SM00181; EGF; 15.

DR SMART; SM00180; EGF\_Lam; 12.

DR SMART; SM00409; IG; 22.

DR SMART; SM00408; IGC2; 21.

DR SMART; SM00406; IGV; 7.

DR SMART; SM00281; Lamb; 3.

DR SMART; SM00282; Lamb; 3.

DR SMART; SM00192; LDLa; 4.

DR SMART; SM00200; SEA; 1.

DR PROSITE; PS00022; EGF\_1; 9.

DR PROSITE; PS01186; EGF\_2; 6.

DR PROSITE; PS00026; EGF\_3; 4.

DR PROSITE; PS00835; IG\_Like; 22.

DR PROSITE; PS50025; LAM\_G\_DOMAIN; 3.

DR PROSITE; PS01248; LAMININ\_TYPE\_EGF; 11.

DR PROSITE; PS01209; LDLRA\_1; 4.

DR PROSITE; PS00668; LDLRA\_2; 4.

DR PROSITE; PS00024; SEA; 1.

KM Signal; Basement membrane; Proteoglycan; Repeat; Glycoprotein;

KM Heparan sulfate; Laminin EGF-like domain; Immunoglobulin domain;

KM Extracellular matrix; EGF-like domain; Disease mutation.

FT SIGNAL 1 21

FT CHAIN 22 4391

FT DOMAIN 80 194

FT DOMAIN 198 235

FT DOMAIN 284 320

FT DOMAIN 324 360

FT DOMAIN 367 404

FT DOMAIN 405 504

FT DOMAIN 521 530

FT DOMAIN 531 730

FT DOMAIN 731 763

FT DOMAIN 764 813

FT DOMAIN 814 871

FT DOMAIN 879 923

BASEMENT MEMBRANE-SPECIFIC HEPARAN SULFATE PROTEOGLYCAN CORE PROTEIN.

SEA.

LDL-RECEPTOR CLASS A 1.

LDL-RECEPTOR CLASS A 2.

LDL-RECEPTOR CLASS A 3.

LDL-RECEPTOR CLASS A 4.

IG-LIKE C2-TYPE 1.

LAMININ EGF-LIKE 1 (N-TERMINAL).

LAMININ DOMAIN IV 1 (DOMAIN III A).

LAMININ EGF-LIKE 1 (C-TERMINAL).

LAMININ EGF-LIKE 2.

LAMININ EGF-LIKE 3.

LAMININ EGF-LIKE 4 (INCOMPLETE).

FT DOMAIN 924 933 LAMININ EGF-LIKE 5 (N-TERMINAL).

FT DOMAIN 934 1125 LAMININ DOMAIN IV 2 (DOMAIN III B).

FT DOMAIN 1126 1158 LAMININ EGF-LIKE 5 (C-TERMINAL).

FT DOMAIN 1159 1208 LAMININ EGF-LIKE 6.

FT DOMAIN 1209 1265 LAMININ EGF-LIKE 7.

FT DOMAIN 1275 1324 LAMININ EGF-LIKE 8.

FT DOMAIN 1325 1334 LAMININ EGF-LIKE 9 (N-TERMINAL).

FT DOMAIN 1335 1529 LAMININ EGF-LIKE 9 (N-TERMINAL).

FT DOMAIN 1530 1562 LAMININ EGF-LIKE 9 (C-TERMINAL).

FT DOMAIN 1563 1612 LAMININ EGF-LIKE 10.

FT DOMAIN 1613 1670 LAMININ EGF-LIKE 11.

FT DOMAIN 1671 1771 LAMININ EGF-LIKE 12.

FT DOMAIN 1772 1865 LAMININ EGF-LIKE 13.

FT DOMAIN 1866 1955 LAMININ EGF-LIKE 14.

FT DOMAIN 1956 2051 LAMININ EGF-LIKE 15.

FT DOMAIN 2052 2151 LAMININ EGF-LIKE 16.

FT DOMAIN 2152 2244 LAMININ EGF-LIKE 17.

FT DOMAIN 2245 2340 LAMININ EGF-LIKE 18.

FT DOMAIN 2341 2436 LAMININ EGF-LIKE 19.

FT DOMAIN 2437 2533 LAMININ EGF-LIKE 20.

FT DOMAIN 2534 2629 LAMININ EGF-LIKE 21.

FT DOMAIN 2630 2726 LAMININ EGF-LIKE 22.

Query Match 5.4%; Score 184; DB 1; Length 4391;

Best Local Similarity 20.4%; Pred. No. 0.0026;

Matches 156; Conservative 97; Mismatches 255; Indels 216; Gaps 37;

5 VPFHLLVQLALIPA-ATQGNKV-VLGGKD-----TVELCTASQKKSIOFW 53

2477 LPARHQVGRRLRLQTPADSGEYVCRVVGSGTQASVLTITQGLSGHSGVAV-- 2534

54 KNSQIILNQGSGFLTKGSKNDKADRSRS--LMDQ--GNP-----LITKL 99

2535 ---PVRISSASLANGHFTLNCVLSQAHPHTTYKKGSGSPSRHQIVGSRRLPQV 2590

100 KIEDSDTYICEVED--OKEEVOLLVFLTANSPTHLIQGQ-STLTLSP--GSSPS 152

2591 TPADSGEYVCRVVGSGTQASVLTITQGLSGHSGVAV--LITKL 2640

153 V-----OC---RSP-----RGNKIQG-----GKTLVSQLELDGSGTCTVLO 188

2641 VVEGQTLNLCVVARQQAIIITWYKRGSLPSRHQTHGSHLRHQMVSADSGEYVCRANN 2700

189 NQKKVEKTIIV-----LAPQASSIVYKKGGEVSEFPPLAFYTEKLTGS 235

2701 NIDALEASIVISVPSAGSPAPSSMPRIEISSSHVAGETLIDN-----CVVQQA 2755

236 GELMQLERASSSSKSWTFPLKQKKEVSKRVTDPKLQMKKPLHLTLPOALPOVAGSG 295

2756 AQTWYKRGSLPSRHQTHGSHLRHQMVSADSGEYVCRANN 2700

296 NLTLALEAKTKLHOENVLVVMTATOLK-----RGRRLRHHVSP-----ADSG 2789

2790 EYVCRWVWGSGPLEASV-LVTIEASGSAHVAPAGAPRIEIPSSSRVABQGTLDLKC 2848

329 EYWGPTSPKMLSLKLEKAKVSKREKPVYTN-----PRAGMQLSLSGOVLSENI 384

2849 VVFGQALAVQWYKRGSLPSRHQTHGSHLRHQMVSADSGEYVCRANN 2700

385 KVLPTWSTPVPAPARPKSKCDKHTHCELLGSGSVLFPKPKDPTLMSITPVTYCVVD 444

2907 LVITEPSSPPIPP-----GLAQPIYIASSSHVTEGQTLIDNLCVVG 2950

445 VSHEDPEVKNNWYVDGEVHNAKTKPREBOYNSTYRVSVLTVLHODMLNGEKYKCVSN 504

2951 QAH---AQTWYKRG-----GSLPARHQVHSGQLHLHVSRA-----DSGEYVCRAS 2995

505 KALP---APLEKTIKAKQOP---REPOVTLTPPSRDELTKNOVSLTCLV-KKPYSDIA 557

2996 GPGEDEAFSTVTVPPSEGSYRLRSFVLSIDPPSSITVQGGQDASFCLIHDAAP--IS 3053

558 VEVESNQOP-ENNYKTPPLVLDSDGFFLXSKLTVDKSRNQGNVSC-----SV 606

DB 3054 LEWKTNQLEEDVNHIS-----NGSI-----ITIVGTRPSNNGTCVAVSNAYGAQSV 3103  
 QY 607 MHEALHNHYTQKSLSLSP-----GLQIDETCAEADGELDLMT 645  
 DB 3104 VNLSVHGPPPTVSVLPGSPVWVKVGAVTLECVAGSPRSARWT 3147

RESULT 60  
 LACI\_MOUSE  
 ID LACI\_MOUSE STANDARD; PRT; 105 AA.  
 AC P01843;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Ig lambda-1 chain C region.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA MEDLINE=83014953; PubMed=6812053;  
 RA Seising E., Miller J., Wilson R., Storb U.;  
 RT "Evolution of mouse immunoglobulin lambda genes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).  
 RN [2]  
 RP SEQUENCE FROM N.A. (MOPC 315).  
 RA MEDLINE=8114806; PubMed=6259534;  
 RA Botheall A.L.M., Paskind M., Schwartz R.C., Sonenshein G.E.,  
 RA Gelfer M.L., Baltimore D.;  
 RT "Dual expression of lambda genes in the MOPC-315 plasmacytoma";  
 RL Nature 290:65-67(1981).  
 RN [3]  
 RP SEQUENCE FROM N.A. (S43).  
 RA MEDLINE=8220143; PubMed=6283385;  
 RA Botheall A.L.M., Paskind M., Rech M., Imanishi-Kari T., Rajewsky K.,  
 RA Baltimore D.;  
 RT "Somatic variants of murine immunoglobulin lambda light chains";  
 RL Nature 298:380-382(1982).  
 RN [4]  
 RP SEQUENCE (MYELOMA PROTEIN MOPC 104E).  
 RA MEDLINE=7110784; PubMed=5276767;  
 RA Appella E.;  
 RT "Amino acid sequences of two mouse immunoglobulin lambda chains";  
 RL Proc. Natl. Acad. Sci. U.S.A. 68:590-594(1971).  
 CC -1- MISCELLANEOUS: The MOPC 315 cell line produces 2 light chains, 1  
 CC normal lambda-2 chain and 1 abnormal lambda-1 chain that is  
 CC missing a large part of the V region. The C region sequence (shown  
 CC here) appears completely normal.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 CC -----  
 CC THIS SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL: J00582; AAAS1636.1; -  
 CC EMBL: J00587; AABS9672.1; -  
 CC PIR: A93922; LIMS.  
 CC PDB: 1JNH; 06-FEB-02.  
 CC InterPro: IPR007110; Ig-like.  
 CC InterPro: IPR003597; Ig cl.  
 CC InterPro: IPR003066; Ig\_MHC.  
 CC Pfam: PF00047; Ig\_1.  
 CC SMART: SM00407; IgC1; 1.  
 CC PROSITE: PS50835; IG\_LIKE; 1.  
 CC PROSITE: PS00290; IG\_MHC; 1.  
 CC Immunoglobulin domain; Immunoglobulin C region; 3D-structure.  
 FT NON\_TER 1 1  
 FT DOMAIN 6 100 IG-LIKE.

FT DISULFID 27 86  
 FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).  
 FT CONFLICT 19 20 ET -> TE (IN REF. 4).  
 FT CONFLICT 56 56 Q -> E (IN REF. 4).  
 FT CONFLICT 75 75 MISSING (IN REF. 4).  
 FT CONFLICT 81 82 HS -> SH (IN REF. 4).  
 FT CONFLICT 85 85 S -> SS (IN REF. 4).  
 FT CONFLICT 96 96 E -> Q (IN REF. 4).  
 SO SEQUENCE 105 AA; 11575 MW; A89P2B09BFCFA018 CRC64;

Query Match 5.2%; Score 177.5; DB 1; Length 105;  
 Best Local Similarity 35.9%; Pred. No. 5.2e-05;  
 Matches 37; Conservative 18; Mismatches 43; Indels 5; Gaps 3;

QY 521 QPR-EPQVYTLPPSDLETKNOVSLTCLVKGPSPDIAVWESNGQP-ENNYKTPPVLD 578  
 DB 1 QKSSPSVTLFPSSSELETNATVCTITDYPGVVYDMKVDGPTQGMETTPSKQ 60  
 QY 579 SDGSPFLYSKLTVDKSRMQGNVFCQVMEALHNHYTQKSL 621  
 DB 61 SNKKWASSYLTUTARAMEHSHSYCOVTHE---GHYVKSLS 100

RESULT 61  
 PGBM\_MOUSE  
 ID PGBM\_MOUSE STANDARD; PRT; 3707 AA.  
 AC Q05783;  
 DT 01-NOV-1995 (Rel. 32, Created)  
 DT 01-NOV-1995 (Rel. 32, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Basement membrane-specific heparan sulfate proteoglycan core  
 DE protein precursor (HSPG) (Perlecan) (PLC).  
 GN HSPG2.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=Melanoma;  
 RA MEDLINE=92078153; PubMed=1744087;  
 RA Noonan D.M., Fulle A., Valente P., Cai S., Horigan E., Sasaki M.,  
 RA Yamada Y., Hassell J.R.;  
 RT "The complete sequence of perlecan, a basement membrane heparan  
 RT sulfate proteoglycan, reveals extensive similarity with laminin A  
 RT chain, low density lipoprotein-receptor, and the neural cell adhesion  
 RT molecule";  
 RL J. Biol. Chem. 266:22939-22947(1991).  
 RN [2]  
 RP SEQUENCE OF 940-1601 AND 1870-2600 FROM N.A., AND PARTIAL SEQUENCE.  
 RA MEDLINE=8904110; PubMed=2972708;  
 RA Noonan D.M., Horigan E.A., Ledbetter S.R., Vogel G., Sasaki M.,  
 RA Yamada Y., Hassell J.R.;  
 RT "Identification of cDNA clones encoding different domains of the  
 RT basement membrane heparan sulfate proteoglycan";  
 RL J. Biol. Chem. 263:16379-16387(1988).  
 CC -1- FUNCTION: This protein is an integral component of basement  
 CC membranes. It is responsible for the fixed negative electrostatic  
 CC charge and is involved in the charge-selective ultrafiltration  
 CC properties. It serves as an attachment substrate for cells.  
 CC -1- SUBUNIT: Purified perlecan has a strong tendency to aggregate in  
 CC dimers or stellate structures. It interacts with other basement  
 CC membrane components such as laminin, prolargin and collagen type  
 CC IV.  
 CC -1- SUBCELLULAR LOCATION: Extracellular.  
 CC -1- TISSUE SPECIFICITY: Found in the basement membrane.  
 CC -1- PTM: CONTAINS THREE HEPARAN SULFATE CHAINS AS WELL AS N-LINKED  
 CC AND O-LINKED OLIGOSACCHARIDES.  
 CC -1- SIMILARITY: Contains 4 LDL-receptor class A domains.  
 CC -1- SIMILARITY: Contains 11 laminin EGF-like domains.  
 CC -1- SIMILARITY: Contains 1 laminin IV domain.  
 CC -1- SIMILARITY: Contains 15 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 3 laminin G-like domains.



```

FT DISULFID 1792 1839 BY SIMILARITY.
FT DISULFID 1886 1932 BY SIMILARITY.
FT DISULFID 1976 2021 BY SIMILARITY.
FT DISULFID 2073 2118 BY SIMILARITY.
FT DISULFID 2170 2215 BY SIMILARITY.
FT DISULFID 2268 2313 BY SIMILARITY.

Query Match 5.2% Score 177.5; DB 1; Length 3707;
Best Local Similarity 17.7%; Pred. No. 0.0055; Index 193; Gaps 25;
Matches 116; Conservative 94; Mismatches 251;

QY 33 KGDVETLTCTASQKSIQFHMKNNSQIKLGNOSFPLTKGPSKLNDRADSRRLMDQGNF 92
1968 EGRTVLVCRAAGVPSASITWRKEG-----GSLFRFQAHGSR----- 2005
QY 93 PLTIKUKLKEDSDTYIC---EVEDQKEVQLIVFGILT-----ANSTTHLL 134
2006 -LRLHMSVADSGEYVCRANNNDIAQETSIMISVSTNSPPAPAPAPRISSSSRYVA 2064
QY 135 QGOSLTLTLESPPGSPSVQCRSPRGKNI-----QGKTLVSQLELDOSGWTCTVNLGN 189
2065 EGGTLNLNCVVPGHAAQVTMHR-KGGSLEPTHHQTHSRRLRYQVSSADSGEYVCSVLSS 2123
QY 190 OKKVEFKIDIVLAFQKASSIVYKKEGEQVEFPLAFTVEKLTGSGELMWQAEASSSK 249
2124 SGPLEASVLSIR-----PAAANV----- 2143
QY 250 SWITFDLKNKEVSKRVATODPKLOMGKLEPLHLTLPOALFOYAGSGLTLAEAKTGKHL 309
2144 -----IPGVVPIRIETSSSRVAEGOTLDLSCVVP-----GQAH 2177
QY 310 QEVNVLVWRATQIQKMLTCVWQPTSPKMLSLKLEKKAHVSKREKPVVNLPEAGMVG 369
2178 AQV-----TMHRGSSLPAGHQVHGMLRLNR-----VSPADSGEYS 2214
QY 370 CLSDSGQVLLNESNIKVLPTWSTPVPPCAPAPKSCDKTHCPCLLGSPVFLPPPKDXT 429
2215 CQVYTGSSGT-LEASVLTITASESPRIAP-----GLAPVYIESSSH 2257
QY 430 LMTSRPEVTCVVDVSHEDPEVKFNMVYDVEVHNAAKTPREBQYNSTYRVSVLTVLH 489
2258 LTEGQVTDLKCVVPGQAH-----AQVTHKRG-----SLPARHQTHSLRLTYOLSPA-- 2306
QY 490 QDMLEKKEVCKXSNKLP---APIETISKAGQP---REPQVYTPSPRDELTKQVS 543
2307 ----DSGEYVQVAGSSHPHEASFKLTVPSQSSFRILSPVISLEPPSTVQOQGDAS 2362
QY 544 LTCLV-KGFYPSDIAVEMESNGQP-ENNYKTPPVLDSDGSFPLY--SKLTVDKSRWQ 598
2363 FKCLIHGAMP--IKYEMKIRDOLEBNVAISP-----NSITTYVAPGPAIMEPTACVA 2415
QY 599 GNVFSC--SYMHEALHNHYTQKSLISP-----GLQIDETCAEADGELDLMT 645
2416 SNVYGMAGQSVNLSVHGPRTVSVLPBEGPVHVKKGKIDTLECISSGEBRSSPRT 2469

RESULT 62
VCAI RAT STANDARD; PRT; 739 AA.
AC P29534;
DT 01-APR-1993 (Rel. 25, Last Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Vascular cell adhesion protein 1 precursor (V-CAM 1).
GN VCAM1 OR VCAM-1.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Lung;
MEDLINE=92181437; PubMed=1371918;

```

```

RA Hession C., Moy P., Tizard R., Chisholm P., Williams C., Wysk M.,
RA Burky L., Miyake K., Kincaid P., Lobb R.;
RT "Cloning of murine and rat vascular cell adhesion molecule-1.";
RL Biochem. Biophys. Res. Commun. 183:163-169(1992).
CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION
CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1
CC INTEGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL
CC TRANSDUCTION. THE VCAM1/VIA4 INTERACTION MAY PLAY A
CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE
CC EMIGRATION TO SITES OF INFLAMMATION.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as
CC well as on macrophage-like and dendritic cell types in both normal
CC and inflamed tissue.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M84488; AAA42332.1; -.
CC PIR; J50675; J50675.
CC HSSP; P19320; 1VCA.
CC InterPro; IPR003987; ICMV VCA-1.
CC InterPro; IPR007110; I9-11k.
CC InterPro; IPR003598; I9 C2.
CC InterPro; IPR003989; VCA-1.
CC Pfam; PF00047; I9; 5.
CC DR PRINTS; PRO1472; ICMVCA-1.
CC DR PRINTS; PRO1474; VCA-1.
CC SMART; SM00408; I9C2; 3.
CC PROSITE; PS50835; I9 LIKE; 5.
CC Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;
CC Repeat; Signal.
CC
CC FT SIGNAL 1 24 PROBABLE.
CC FT CHAIN 25 739 VASCULAR CELL ADHESION PROTEIN 1.
CC FT DOMAIN 25 698 EXTRACELLULAR (POTENTIAL).
CC FT TRANSMEM 699 720 POTENTIAL.
CC FT DOMAIN 721 739 CYTOPLASMIC (POTENTIAL).
CC FT DOMAIN 25 111 IG-LIKE C2-TYPE 1.
CC FT DOMAIN 119 212 IG-LIKE C2-TYPE 2.
CC FT DOMAIN 223 309 IG-LIKE C2-TYPE 3.
CC FT DOMAIN 312 397 IG-LIKE C2-TYPE 4.
CC FT DOMAIN 408 506 IG-LIKE C2-TYPE 5.
CC FT DOMAIN 514 595 IG-LIKE C2-TYPE 6.
CC FT DOMAIN 601 682 IG-LIKE C2-TYPE 7.
CC FT DISULFID 47 95 BY SIMILARITY.
CC FT DISULFID 52 99 BY SIMILARITY.
CC FT DISULFID 137 195 BY SIMILARITY.
CC FT CARBOHYD 273 273 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 424 424 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 531 531 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 561 561 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL).
CC SQ SEQUENCE 739 AA; 81246 MW; 5C608E5A1A1B100C CRC64;

Query Match 5.1% Score 174.5; DB 1; Length 739;
Best Local Similarity 20.4%; Pred. No. 0.001; Index 185; Gaps 29;
Matches 125; Conservative 81; Mismatches 185;

QY 33 KGDVETLTCTASQKSIQFHMKNNSQIKLGNOSFPLTKGPSKLNDRADSRRLMDQGNF 92
238 EGAAYVMTCASGSLPAPELFMSK-----LDNGVQLL-----SGNA 274
QY 93 PLTIKUKLKEDSDTYIC---EVEDQKEVQLIVFGILT-----ANSTTHLL 134
275 TLTLLMRMDSGIYVCEGVNLVGRDKTEVELVQKRPFTVDISPSQVAAVAGDSVLT 334
QY 143 LSPSPGSSPSVQCRSPRGKNIQ-----GKTLVSQLELDOSGWTCTVNLQNKVVF 195

```

[illegible]

CC	-1- SIMILARITY: Contains 1 DBL-homology (DH) domain.
CC	-1- SIMILARITY: Contains 1 fibronectin type III domain.
CC	-1- SIMILARITY: Contains 49 immunoglobulin-like C2-type domains.
CC	-1- SIMILARITY: Contains 1 PH domain.
CC	-1- SIMILARITY: Contains 5 RCSD domains.
CC	-1- SIMILARITY: Contains 1 SH3 domain.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/or_send_an_email_to_license@isb-sib.ch">http://www.isb-sib.ch/announce/</a>
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).
CC	-----
DR	EMBL, U33058; AAB0542.1; -.
DR	EMBL, AF003131; AAB54132.2; -.
DR	PDB, 1FHO; 20-DEC-00.
DR	WormPep, C09D1.1; CB30426.
DR	InterPro, IPR008957; FN_III-like.
DR	InterPro, IPR003961; FN_III.
DR	InterPro, IPR007110; Ig-like.
DR	InterPro, IPR003598; Ig C2.
DR	InterPro, IPR003066; Ig_MHC.
DR	InterPro, IPR001849; PH.
DR	InterPro, IPR007850; RCSD.
DR	InterPro, IPR000219; RhogEF.
DR	InterPro, IPR001452; SH3.
DR	Pfam, PF000041; fn3; 1.
DR	Pfam, PF000047; Ig; 47.
DR	Pfam, PF00169; PH; 1.
DR	Pfam, PF05177; RCSD; 5.
DR	Pfam, PF00621; RhogEF; 1.
DR	Pfam, PF00018; SH3; 1.
DR	SMART, SMO03408; IgC2; 23.
DR	SMART, SMO0325; RhogEF; 1.
DR	SMART, SMO0326; SH3; 1.
DR	PROSITE, PSS0010; DH 2; 1.
DR	PROSITE, PSS0835; IG_LIKE; 49.
DR	PROSITE, PSS0003; PH_DOMAIN; 1.
DR	PROSITE, PSS0002; SH3; 1.
KM	Muscle protein; Immunoglobulin domain; Repeat; SH3 domain;
KM	3D-structure.
FT	DOMAIN 63 127 SH3.
FT	DOMAIN 152 330 DH.
FT	DOMAIN 342 498 PH.
FT	DOMAIN 547 633 IG-LIKE C2-TYPE 1.
FT	DOMAIN 648 736 IG-LIKE C2-TYPE 2.
FT	DOMAIN 748 838 IG-LIKE C2-TYPE 3.
FT	DOMAIN 946 1033 IG-LIKE C2-TYPE 4.
FT	DOMAIN 1044 1132 IG-LIKE C2-TYPE 5.
FT	DOMAIN 1140 1227 THR-RICH.
FT	DOMAIN 1272 1315 RCSD 1.
FT	DOMAIN 1375 1475 RCSD 2.
FT	DOMAIN 1479 1585 RCSD 3.
FT	DOMAIN 1597 1695 RCSD 4.
FT	DOMAIN 1700 1799 RCSD 5.
FT	DOMAIN 1800 1860 RCSD 4.
FT	DOMAIN 1982 2067 IG-LIKE C2-TYPE 7.
FT	DOMAIN 2071 2163 IG-LIKE C2-TYPE 8.
FT	DOMAIN 2171 2261 IG-LIKE C2-TYPE 9.
FT	DOMAIN 2269 2359 IG-LIKE C2-TYPE 10.
FT	DOMAIN 2367 2455 IG-LIKE C2-TYPE 11.
FT	DOMAIN 2463 2564 IG-LIKE C2-TYPE 12.
FT	DOMAIN 2563 2651 IG-LIKE C2-TYPE 13.
FT	DOMAIN 2657 2746 IG-LIKE C2-TYPE 14.
FT	DOMAIN 2754 2858 IG-LIKE C2-TYPE 15.
FT	DOMAIN 2887 2980 IG-LIKE C2-TYPE 16.
FT	DOMAIN 2994 3081 IG-LIKE C2-TYPE 17.
FT	DOMAIN 3087 3183 IG-LIKE C2-TYPE 18.
FT	DOMAIN 3189 3280 IG-LIKE C2-TYPE 19.
FT	DOMAIN 3286 3376 IG-LIKE C2-TYPE 20.
FT	DOMAIN 3384 3469 IG-LIKE C2-TYPE 21.

```
FT DOMAIN 3482 3572 IG-LIKE C2-TYPE 22.
FT DOMAIN 3580 3667 IG-LIKE C2-TYPE 23.
FT DOMAIN 3686 3777 IG-LIKE C2-TYPE 24.
FT DOMAIN 3817 3908 IG-LIKE C2-TYPE 25.
FT DOMAIN 3920 4009 IG-LIKE C2-TYPE 26.
FT DOMAIN 4018 4106 IG-LIKE C2-TYPE 27.
FT DOMAIN 4109 4201 IG-LIKE C2-TYPE 28.
FT DOMAIN 4212 4297 IG-LIKE C2-TYPE 29.
FT DOMAIN 4302 4387 IG-LIKE C2-TYPE 30.
FT DOMAIN 4400 4485 IG-LIKE C2-TYPE 31.
FT DOMAIN 4489 4580 IG-LIKE C2-TYPE 32.
FT DOMAIN 4678 4771 IG-LIKE C2-TYPE 33.
FT DOMAIN 4881 4961 IG-LIKE C2-TYPE 34.
FT DOMAIN 4873 4961 IG-LIKE C2-TYPE 35.
FT DOMAIN 4965 5057 IG-LIKE C2-TYPE 36.
FT DOMAIN 5067 5160 IG-LIKE C2-TYPE 37.
FT DOMAIN 5171 5260 IG-LIKE C2-TYPE 38.
FT DOMAIN 5277 5366 IG-LIKE C2-TYPE 39.
FT DOMAIN 5383 5472 IG-LIKE C2-TYPE 40.
FT DOMAIN 5487 5578 IG-LIKE C2-TYPE 41.
FT DOMAIN 5595 5685 IG-LIKE C2-TYPE 42.
FT DOMAIN 5701 5790 IG-LIKE C2-TYPE 43.
FT DOMAIN 5815 5904 IG-LIKE C2-TYPE 44.
FT DOMAIN 5925 6014 IG-LIKE C2-TYPE 45.
FT DOMAIN 6038 6130 IG-LIKE C2-TYPE 46.
FT DOMAIN 6150 6239 IG-LIKE C2-TYPE 47.
FT DOMAIN 6275 6368 FIBROECTIN TYPE-III.
FT DOMAIN 6413 6502 IG-LIKE C2-TYPE 48.
FT DOMAIN 6507 6596 IG-LIKE C2-TYPE 49.
FT DISULFID 568 621 POTENTIAL.
FT DISULFID 2908 2975 POTENTIAL.
FT DISULFID 3015 3065 POTENTIAL.
FT DISULFID 3707 3759 POTENTIAL.
FT DISULFID 3826 3890 POTENTIAL.
FT DISULFID 5092 5157 POTENTIAL.
FT DISULFID 5298 5350 POTENTIAL.
FT DISULFID 5508 5560 POTENTIAL.
FT DISULFID 5616 5669 POTENTIAL.
FT DISULFID 5722 5764 POTENTIAL.
FT DISULFID 5836 5901 POTENTIAL.
FT DISULFID 5946 5998 POTENTIAL.
FT DISULFID 6036 6171 POTENTIAL.
FT DISULFID 6421 6486 POTENTIAL.
FT DISULFID 2137 2137 A -> P (IN REF. 1).
FT CONFLICT 2245 2247 AKA -> PKP (IN REF. 1).
FT CONFLICT 2258 2258 A -> P (IN REF. 1).
FT CONFLICT 2284 2284 E -> G (IN REF. 1).
FT CONFLICT 2297 2297 M -> I (IN REF. 1).
FT CONFLICT 3531 3531 A -> G (IN REF. 1).
FT CONFLICT 3884 3888 DAGEV -> RRRRI (IN REF. 1).
FT CONFLICT 3929 3929 A -> V (IN REF. 1).
FT CONFLICT 5134 5134 A -> P (IN REF. 1).
FT CONFLICT 5145 5145 T -> S (IN REF. 1).
FT CONFLICT 5185 5185 G -> A (IN REF. 1).
FT CONFLICT 5199 5199 K -> N (IN REF. 1).
FT CONFLICT 5202 5202 L -> F (IN REF. 1).
FT CONFLICT 5213 5213 F -> L (IN REF. 1).
FT CONFLICT 6178 6178 A -> G (IN REF. 1).
FT CONFLICT 6268 6268 K -> E (IN REF. 1).
SQ SEQUENCE 6632 AA; 731665 MW; 262d3bDD62960E89 CRC64;

Query Match 5.1%; Score 174; DB 1; Length 6632;
Best Local Similarity 20.7%; Pred. No. 0.02; Indels 170; Gaps 24;
Matches 132; Conservative 82; Mismatches 255;

QY 20 PAATQGNKVVIGKGGDTVELCTASOKKSIQFHWKNSNQIKILGNQGSFLTGPKSLNDR 79
Db 3580 PLTVQGIKPYEVGEVQKRAELVNVGEKPEPEVKMFQDGVIALDNGVIEKKG----- 3632
QY 80 ADSRRSLMDQGNFPLIKIKLIEDSDTYICEVEDQ-----KEEV 118
Db 3633 -----ENGSHTLVTKDNNADFGKTCQATNKAKGDETVGLKIPKYSFEKQTAEEV 3684
```

```
QY 119 QLVVLGTANSDPHLLOGSLTLTLESPGSSPSVQCRSPRGKNIQGGKTLSSOLELOD 178
Db 3685 KQLF--IEPLKETPAVEGDTVLVECKANKESHPIQK-----PKNQPVISIGMHQLEIVE 3738
QY 179 SGTWTCTVLQNGKXKVEFKIDIVLAFQKASSIVYKKEGQVEFSPLAFTVEKLTGSGEL 238
Db 3739 DGNIKLTI-QNAK-----EDVG-----AVRCBAVNWAGK- 3767
QY 239 WQAERASSKSMITFDLKNKEVSXKRVLTQDPKLOMGKLPPLHLT-----LPQALPOYAS 294
Db 3768 -----ANTNADKTIQPAKVEHVTDESQLEBIGQFETV 3802
QY 295 GNLTALTEAKTGKHAQEVNLVVMRATOLQK---LTCEVWGPSTPMLSLTENKAKV 351
Db 3803 GD-TASSKTDTRGAPFELFELLRSCTVTEKQAAILKCKVKGRPRRIKWT--KEGEVEM 3859
QY 352 SRRKRPW-----VLNPAQMWQC-LYSDSGVLLSNIKVLPTMSTPVPCPA 398
Db 3860 SARVAEAHKDQGLTLTFPDNVTQADAGEYRCBAENEYGSAMTEGPITVLEGAPKIDGBA 3919
QY 399 PEPKSCDKHTHCPBELLGGSVFLFPKPKDITLMISRTPEVTCVVVDVSHEDPEVKENWYV 458
Db 3920 PD-----FLQPVKPAV-----TVGETVAVLEGKISGKRPVSKYK 3955
QY 459 DQVEVHNAKTKPREQVNSTYRVSVLTVLHQDWLNGKEYKCKVSNKALPAPIEKTISKY 518
Db 3956 NEBELKPSRVAVIENLDQTOR---LVTNKLKDDMDYRCEASNEPFDVSDVLT--- 4009
QY 519 KQAPRPQVYTLTPSRDELTKNQVSLT-----CLVKGFPYSDIAVESNQCPENNYKT 572
Db 4010 ---VKEP-AQVAPGFPEKLSAIQVKETETAKFECKVSGTKPD---VKMFQDGPLKEDKR 4062
QY 573 TPPLVDSGSPFLYSGKLTVDKSRM-QQGNVFCSSVWHEA 610
Db 4063 VHESTDDT---QLVIEDSKTDQGN-YRIEVSND 4096

RESULT 64
LAC2_RAT
ID LAC2_RAT STANDARD; PRT; 104 AA.
AC P20767;
DT 01-FEB-1991 (rel. 17, Created)
DT 01-FEB-1991 (rel. 17, Last sequence update)
DT 10-OCT-2003 (rel. 42, Last annotation update)
DE Ig lambda-2 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_Taxid=10116;
RN [1]
RX MEDLINE=87305594; PubMed=3114047;
RT Steen M.L., Hellman U., Petersson U.;
RL "The immunoglobulin lambda locus in rat consists of two C lambda
genes and a single V lambda gene."
RT Gene 55:75-84(1987).
CC -!- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M2521; AAA1420.1; ALT_INIT.
CC HSSP; P01842; 2MCG.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_C1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1.
DR SMART; SM00407; IgC1.1.
```



```

FT  VARIANT 302 302 F -> L.
FT  VARIANT 316 316 F -> L.
FT  VARIANT 337 337 G -> R.
FT  VARIANT 367 367 S -> N.
FT  VARIANT 422 422 Q -> L.
FT  VARIANT 429 429 I -> F.
FT  VARIANT 433 433 D -> E.
SQ  SEQUENCE 506 AA; 55093 MM; 6B7E310677FC9CB CRC64;

Query Match 5.0%; Score 171, DB 1; Length 506;
Best Local Similarity 24.4%; Pred. No. 0.0011;
Matches 65; Conservative 40; Mismatches 115; Indels 46; Gaps 9;

QY 400 EPKSCDKTH-----TCPELLGGPSVFLFPPKPKDLMISRTEGVGVVDVSH--EDPE 451
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 133 EFKSGEPTLTVSAKSPPLSGPTV-----RATPEQVNTFTCTISHGSPRN 179
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 452 VKENMYDVGEVHNNAKTPREEOYNSTRVSVLTVLHODMLNGEKYCKVSKALPA-- 509
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 180 ISLKMFKNGNELSASQTSVPDENNNYSINSTTKYLLATGDVHSQVIGCAVAVTLQGGP 239
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 510 PIKRTISKAKGQPREPOVYTLPPSRDELTL-----KNQVSLTCLVKGFPESDIWESNG 564
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 240 PL-----RGTANLSETIRVPTL-EITGSPAGNQVNTCCVNNKFFPRHLQTLWLENG 291
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 565 QPENNYKTPPVLDSDGSFFLYGKLTVDKSRMGGQNVFSGSVWHEALHNHYTKSLSLSP 624
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 292 NMSRTEASVFEVKDGTFTQTSFELVSSAHRBAVVLTCVSHDG-----OPAVSKN 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 625 GLQLDTCARADQDEL----DGLMTT 646
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 346 TLEVSAPQKDDGTGQTFGRPNDSNMTS 371
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 66
CEAS HUMAN STANDARD; PRT; 702 AA.
ID CEAS_HUMAN
AC P06731;
DT 01-JUN-1988 (Rel. 06, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Carcinoembryonic antigen-related cell adhesion molecule 5 precursor
DE Carcinoembryonic antigen (CEA) (Meconium antigen 100) (CD66e
DE antigen)
GN CEACAM5 OR CEA.
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OC NCBI_TaxID=9606;
OX [1]
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90258861; PubMed=2342461;
RA Schrewe H., Thompson J., Bona M., Hefta L.J.F., Maruya A.,
RA Hassauer M., Shively J.E., von Kleist S., Zimmermann W.;
RT "Cloning of the complete gene for carcinoembryonic antigen: analysis
RT of its promoter indicates a region conveying cell type-specific
RT expression.";
RL Mol. Cell. Biol. 10:2738-2748(1990).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=88038876; PubMed=3670312;
RA Beaubien N., Benchinol S., Cournoyer D., Fuks A., Stanners C.P.;
RT "Isolation and characterization of full-length functional cDNA clones
RT for human carcinoembryonic antigen.";
RL Mol. Cell. Biol. 7:3221-3230(1987).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=89122014; PubMed=3220478;
RA Barnett T., Goebel S.V., Notdurft M.A., Elting J.J.;
RT "Carcinoembryonic antigen family: characterization of cDNAs coding
RT for CEA and CEA and suggestion of nonrandom sequence variation in
RT their conserved loop-domains.";
RL Genomics 3:59-66(1988).

```

```

RN [4]
RP SEQUENCE OF 5-702 FROM N.A.
RX MEDLINE=87128144; PubMed=3814146;
RA Oikawa S., Nakazato H., Kosaki G.;
RT "Primary structure of human carcinoembryonic antigen (CEA) deduced
RT from cDNA sequence.";
RL Biochem. Biophys. Res. Commun. 142:511-518(1987).
RN [5]
RP SEQUENCE OF 331-702 FROM N.A.
RX MEDLINE=87204247; PubMed=3033671;
RA Zimmermann W., Ortlieb B., Friedrich R., von Kleist S.;
RT "Isolation and characterization of cDNA clones encoding the human
RT carcinoembryonic antigen reveal a highly conserved repeating
RT structure.";
RL Proc. Natl. Acad. Sci. U.S.A. 84:2960-2964(1987).
CC -1- SUBCELLULAR LOCATION: Attached to the membrane by a GPI-anchor.
CC -1- TISSUE SPECIFICITY: Found in adenocarcinomas of endoderally
CC derived digestive system epithelium and fetal colon.
CC -1- PTM: COMPLEX IMMUNOREACTIVE GLYCOPROTEIN WITH A MW OF 180 KDa
CC COMPRISING 60% CARBOHYDRATE.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. CEA family.
CC -1- SIMILARITY: Contains 7 immunoglobulin-like domains.
CC -1- DATABASE: NAME=PROW; NOTE=CD guide CD66e entry;
CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd66e.htm".
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, M17303; AAB59513.1; -; ALT_SEQ.
DR EMBL, M59262; AAA62835.1; -; JOINED.
DR EMBL, M59255; AAA62835.1; -; JOINED.
DR EMBL, M59256; AAA62835.1; -; JOINED.
DR EMBL, M59257; AAA62835.1; -; JOINED.
DR EMBL, M59258; AAA62835.1; -; JOINED.
DR EMBL, M59259; AAA62835.1; -; JOINED.
DR EMBL, M59260; AAA62835.1; -; JOINED.
DR EMBL, M59261; AAA62835.1; -; JOINED.
DR EMBL, M59709; -; NOT_ANNOTATED_CDS.
DR EMBL, M59710; -; NOT_ANNOTATED_CDS.
DR EMBL, M29540; AAA51957.1; -;
DR EMBL, X16455; CAA34747.1; -;
DR EMBL, M15042; AAA51963.1; -;
DR EMBL, M16234; AAA51972.1; -;
DR PIR, A36319; A36319.
DR PDB, 1B07; 04-JUL-00.
DR GeneW; HGNC:1817; CEACAM5.
DR MIM, 114890; -.
DR GO; GO:0005887; C:integral to plasma membrane; TAS.
DR InterPro; IPR007110; IG-like.
DR Pfam; PF00047; Ig_6.
DR PROSITE; PS50835; IG_LIKE; 6.
KW Immunoglobulin domain, Glycoprotein, Lipoprotein, GPI-anchor;
KW Membrane; Signal; Repeat; 3D-structure.
FT SIGNAL 1 34
FT CHAIN 35 685
FT PROPEP 686 702
FT DOMAIN 35 144
FT DOMAIN 146 237
FT DOMAIN 238 322
FT DOMAIN 324 415
FT DOMAIN 416 498
FT DOMAIN 502 593
FT DOMAIN 594 677
FT LIPID 685 685
FT CARBOHYD 104 104
FT CARBOHYD 115 115
FT CARBOHYD 152 152

```



```

FT CARBOHYD 182 182 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 197 197 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 204 204 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 208 208 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 256 256 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 274 274 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 288 288 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 292 292 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 309 309 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 330 330 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 351 351 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 360 360 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 432 432 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 466 466 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 480 480 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 508 508 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 529 529 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 553 553 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 560 560 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 580 580 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 612 612 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 650 650 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CARBOHYD 665 665 N-LINKED (GLCNAC. . .) (POTENTIAL)
FT CONFLICT 320 320 MISSING (IN REF. 4)
SQ SEQUENCE 702 AA; 76795 MW; 6299AE26C0DBB5C CRC64;

```

```

Query Match 4.8%; Score 165; DB 1; Length 702;
Best Local Similarity 17.6%; Pred. No. 0.0039;
Matches 118; Conservative 109; Mismatches 255; Indels 190; Gaps 31;

```

```

QY 20 PATQGNKRVNKGKDDVETLCTASQKSIQFHWKNSQIKILNQSGFLTKGSKLND 79
DB 147 PSISSNNSKPEVDK-DAVAFCEPTQ2ATYLMWVNSLPV----- 187
QY 80 ADSRRSLWDGNFPLIKLKIEDSDTYICEVD-----QKEVOL-LVFGILTANS--- 129
DB 188 --SRLQSGNRTLTLEFNTVRNDPTASYKCEFNQNVASRSDSVILNLYGPDAPPTISPL 245
QY 130 DTHLQGSLLTLESPPGSSPVQCR-SPRGKIQGSKTSLVSQLEIQDSGCTWTCVYLQ 188
DB 246 NTSYRSGENLMSCHA--ASNPPAOYSFVNGTFOQSTQELFIPITVNNSGSYTCQAHN 303
QY 189 NQKRV-EFKIDIVLAFQKASSIYKKGGEVSEFSPLATVE-KLNGSGSLMQAERAS 246
DB 304 SDTGILNRTTITTYIYAEPPKPFITSNNSNPEDBDALTCPELIONTYLWM----- 357
QY 247 SSKSWITFDLKNKEVSVKRVTDPELQMGKLLPLHLTPQALPOYAGSGNLTLLAEATG 306
DB 358 -----VANQSLPV-----SPRLQ-----SNDNTLTLLSVTR 385
QY 307 KLHOBVNLVNRATQLOKNTCE-----VWGPTSP-----KLMLSLKL 344
DB 386 -----NDVGPEYECIGNELSVDSHSDPVILNVLVGPDPPTISPYTYRRPGVNLISCHA 439
QY 345 ENK-----EAKVSKREKPVWVLN---PEAGMQCLDSDSQVLESIKVLPMTST 392
DB 440 ASNPPAOYSWLDENIQHTDELFSNTEKNSGLYTQANNSSAGHSRTTVKTI-TVSA 498
QY 393 PVPCPAPRPSKCDKHTHCPPELLGGPSVFLPPPKPDOTLMSRTEPEVTCVVVDVSHEDREV 452
DB 499 ELPRPSISSNNSK-----FVEDKDAVAFCEBA-----ONT 530
QY 453 KFNWYVDGVEVHNAKTKPREQVNSTYRVSVLTVLHODWLNKGEYKCKVSNKALPAPIE 512
DB 531 TYLMWVNG---QSLPVSPRLQLSNGN-RTLTLEFNTVRND--ARAAYVGIGN----- 575
QY 513 KTISAKQGPREFQV-----YTLPSRDELTKNOVSLTCLVKGFTYPSDIAVWESNGQ 565
DB 576 -SVSNRSDPVTLDVLPDTPPIISPPSSYLSGANLWSC-HSASNSP-QYSWRINGI 632
QY 566 PENNYKTPPVLDSDGSFFLYSKLTVDSKRWQGVFSQVMHEAL-HNHYTKSLSL- 622

```

```

DB 633 PQOHTQV-----LPIAKITPNMN-----GTACFVSNLATGRNNSIVKSTVSA 676
QY 623 ---SPGLQDET 631
DB 677 SGTSFGLSAGAT 688

```

## RESULT 67

ID\_LAC3\_MOUSE STANDARD; PRT; 104 AA.

AC P01845;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Ig lambda-3 chain C region.

OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_Taxid=10090;

[1]  
 RN SEQUENCE FROM N.A.

RX MEDLINE=83014953; PubMed=6812053;  
 RA Seising E., Miller J., Wilson R., Scorb U.;  
 RT "Evolution of mouse immunoglobulin lambda genes";  
 RL Proc. Natl. Acad. Sci. U.S.A. 79:4681-4685(1982).

[2]  
 RN SEQUENCE OF 1-18 (MYELOMA PROTEIN CBPC-49).

RA Breyer R.M., Sauer R.T., Eisen H.N.;  
 RT "The variable region of mouse lambda-3 chains";  
 RL ICN UCLA Symp. Mol. Cell. Biol. 20:105-110(1981).

[3]  
 RN SEQUENCE OF 10-104 (MYELOMA PROTEIN CBPC-49 AND MONOCLONAL ANTIBODY

RP 8-47).

RX MEDLINE=8123782; PubMed=6165998;  
 RA Azuma T., Steiner L.A., Eisen H.N.;

RT "Identification of a third type of lambda light chain in mouse  
 immunoglobulins";

RL Proc. Natl. Acad. Sci. U.S.A. 78:569-573(1981).

CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC EMBL; J00585; AAB59670.1; -

DR PIR; B93922; L3MS.

DR HSBP; P01842; 2MCG.

DR InterPro; IPR007110; Ig-like.

DR InterPro; IPR003597; Ig\_C1.

DR InterPro; IPR003006; Ig\_MHC.

DR Pfam; PF00047; Ig\_1.

DR SMART; SM00407; Igcl, 1.

DR PROSITE; PS00835; Ig\_LIKE, 1.

DR PROSITE; PS00290; IG\_MHC, 1.

KW Immunoglobulin domain; Immunoglobulin C region.

FT NON TER 1 1

FT DOMAIN 6 99

FT DISULFID 27 85

FT DISULFID 103 103

SQ SEQUENCE 104 AA; 11371 MW; 83CEBCD4AA348EF1 CRC64;

Query Match 4.8%; Score 163; DB 1; Length 104;

Best Local Similarity 34.3%; Pred. No. 0.0043;

Matches 35; Conservative 16; Mismatches 47; Indels 4; Gaps 2;

QY 521 QPRE-POVYTLPPSRDELTKNOVSLTCLVKGFTYPSDIAVWESNGQENNYKTPPVLD 579

DB 1 QPKSTPLTWMPSPBELQENKATLVCLISNFSGVTVMKANGFTITGQVDTSNPTKE 60



```

CC -1- ALTERNATIVE PRODUCTS:
CC Event=Alternative splicing; Named isoforms=5;
CC Comment=Additional isoforms seem to exist;
CC Name=1;
CC IsoId=p97798-1; Sequence=Displayed;
CC Name=2;
CC IsoId=p97798-2; Sequence=VSP_002594;
CC Name=3;
CC IsoId=p97798-3; Sequence=VSP_002595;
CC Name=4;
CC Note=Expression developmentally regulated;
CC Note=Expression developmentally regulated;
CC Name=5;
CC IsoId=p97798-5; Sequence=VSP_002597;
CC Note=Expression developmentally regulated;
CC TISSUE SPECIFICITY: Widely expressed.
CC -1- DEVELOPMENTAL STAGE: EXPRESSED UNIDIRECTIONALLY THROUGHOUT THE MID TO LATE STAGES OF GESTATION AND IN ADULT TISSUES. STRONG EXPRESSION IS OBSERVED IN THE VENTRAL REGION OF THE VENTRICULAR ZONE OF THE E15.5 MOUSE NEURAL TUBE, AS WELL AS IN THE VENTRICULAR ZONES OF THE MESENCEPHALON AND RHOMBENCEPHALON. ISOFORMS 3 AND 4 ARE EXPRESSED AT HIGHER LEVEL COMPARED TO OTHER ISOFORMS BETWEEN E11.5 AND E16.5.
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily, DCC family.
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 6 fibronectin type III domains.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see http://www.isb-sib.ch/announce/ or send an email to license@db-sib.ch).
CC -----
DR EMBL; Y09535; CAAT70727.1; -.
DR HSSP; P02751; ITTF.
DR MGD; MGI:1097159; Neol.
DR Interpro; IPR008957; FN_III-like.
DR Interpro; IPR003961; FN_III.
DR Interpro; IPR003962; FNIII subd.
DR Interpro; IPR007110; Ig-like.
DR Interpro; IPR003598; Ig_C2.
DR Pfam; PF00047; fn3; 6.
DR Pfam; PF00047; Ig; 4.
DR PRINTS; PR00014; FN3YPEIII.
DR SMART; SM00060; FN3; 6.
DR SMART; SM00408; IGC2; 4.
DR PROSITE; PS50835; IG_LIKE; 4.
KW Cell adhesion; Repeat; Signal; Transmembrane; Immunoglobulin domain; Glycoprotein; Alternative splicing.
KW Glycoprotein; Alternative splicing.
FT SIGNAL 1 36
FT CHAIN 37 1493
FT DOMAIN 37 1136
FT TRANSMEM 1137 1157
FT DOMAIN 1158 1493
FT DOMAIN 63 158
FT DOMAIN 163 249
FT DOMAIN 254 347
FT DOMAIN 352 437
FT DOMAIN 467 564
FT DOMAIN 567 660
FT DOMAIN 661 760
FT DOMAIN 766 860
FT DOMAIN 881 981
FT DOMAIN 982 1083
FT DOMAIN 1149 1153
FT DISULFID 85 140
FT DISULFID 184 232
FT DISULFID 281 331
FT DISULFID 373 421
FT CARBOHYD 84 84
FT N-LINKED (GLCNAC... ) (POTENTIAL).
FT EXTRACELLULAR (POTENTIAL).
FT POTENTIAL.
FT CYTOPLASMIC (POTENTIAL).
FT IG-LIKE C2-TYPE 1.
FT IG-LIKE C2-TYPE 2.
FT IG-LIKE C2-TYPE 3.
FT IG-LIKE C2-TYPE 4.
FT FIBRONECTIN TYPE-III 1.
FT FIBRONECTIN TYPE-III 2.
FT FIBRONECTIN TYPE-III 3.
FT FIBRONECTIN TYPE-III 4.
FT FIBRONECTIN TYPE-III 5.
FT FIBRONECTIN TYPE-III 6.
FT POLY-VAL.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT BY SIMILARITY.
FT N-LINKED (GLCNAC... ) (POTENTIAL).

```

Query Match	Best Local Similarity	4.8%; Score 162.5; DB 1; Length 1493;	
Matches 159; Conservative	21.8%; Pred. No. 0.015;		
	79; Mismatches 282; Indels 211; Gaps		
FT CARBOHYD	221	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	337	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	501	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	520	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	670	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	746	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT CARBOHYD	940	N-LINKED (GLCNAC . . .)	(POTENTIAL)
FT VASPPLIC	442	Missing (In isoform 2).	
FT VASPPLIC	863	/FtId=VSP_002594.	
FT VASPPLIC	878	Missing (In isoform 3).	
FT VASPPLIC	1086	/FtId=VSP_002595.	
FT VASPPLIC	1096	Missing (In isoform 4).	
FT VASPPLIC	1279	/FtId=VSP_002596.	
FT VASPPLIC	1331	Missing (In isoform 5).	
FT VASPPLIC	1493	/FtId=VSP_002597.	
FT VASPPLIC	163159	MM; 44IDE919DSE17C0E CRC64;	
Query Match	4.8%; Score 162.5; DB 1; Length 1493;		
Best Local Similarity	21.8%; Pred. No. 0.015;		
Matches 159; Conservative	79; Mismatches 282; Indels 211; Gaps		
10 LLLVLTALL--PAA-----	TOGNKY-----	VLKGGDTVELT	40
25 LLLLLPLLLLLLGRASGAATKSGPRRSGASVRTFFFLVEPYDTLSEVRGSSVILL	84		
41 CTASQKKSIOFHMKNSNQIKLNGQSFSLTGSPKLNDRASRSLSIMQGFPLIKYL	99		
85 CSAVSESPNTEWK-----	XDGF-----	LINESDRKQLPDGS--LFISNV	126
100 ----KIEDSDYIK--EVED-----	QKEVQLVFEGL----	TANSDFHLLQGSLLTLDES	145
127 HSKNKNDDEGYQCVATYDNLGTVSRAKLTVAGLPRFTSQPPESSVYVNSALINEV	186		
146 PGSSPSV--QCRSP-----	RKNIOGKTLVSQLELDQSDGWTCTVLON----	OKK	192
187 NADLVPEFRWEQNQPLLLDRIVKLTSPSG--TLVISNMTGEGGGLVRCIVESGCPKPSDE	245		
193 VEFKI-----DIVVLAFOKASSIYYKKRGEVPSFLPAFVEKLTGSGELMWQBRASS	247		
246 AELKVLQDPEITVVLVFLMRSSMKKTGQ--SAILPCVSGLPAPVVR-----	293		
248 SKSWITFDLKNKEVSVKRVITODPKLQNGKPLHLTLPLQALPOY--AGSGLT-----	298		
294 -----MKNEVELDRESSGRLVILAGGCLGISDVEDDAGTFCIADNGNKTVQAQAB	345		
299 LALFAKTGKLHGEVNLVVMRATQLOKULTGCVWGPTSPKML-----SLKENKEAKSVKR	354		
346 LTVQVPPGFLLQPNANIYAHESMDIV--FECEVTGKPTPYKVMKNGDVVIPSDFKIYE	403		
355 E--KPWWVLNPEAMMOCLL--SDSGQ-----VLTSNINIVLPTWSTPRVCPAPKPSG	404		
404 HNLQVLGVNSDEGFTQCIANDVGNNAQGLIILEHDV--AIPF-----LPPSLTSA	456		
405 DKHTTCBELLGSPVFLPPPKXDTLMSRTEVTCVVVDVSHEDPEVKFMYYDVGEVH	464		
457 TTDHLAATGTP-----LPSAPRD-----VVASLVSTRF--IKLFWRTAPSDPH	498		
465 NAKTKPREEOYNSTRVVSULTVLHODMLNGEKCKVSKNALPAPLEKITS-----	516		
499 GD-----NLTVSVFYTKEGVDR-----RVENTSQGEBOVTTIQNIMPATVY	540		
517 --KAKGQPREQVYTLTPSRDELTKNOVSL-----TCLVKGFYSDDIAVEMES-----NGO	565		
541 IFKMAQNKHSGSSAPLRYE--IQPBYQLPGRANIRAYATSPSTIVYVTFETPLSGNGE	599		
566 PENNYKTPPVLDSDGSEFLYSKLTVDKSRWQGNVFSQVMBALHNHYTQKSLT---	622		
600 IQ--NYK-----LTYMEKGTJKE--QDIDVSHSYTINGL--KKYTEYFRRVAY	643		
623 ---SPGLQDLE 630			
644 NKHGPGVSTOD 654			

RESULT 70  
 ID KMLS\_CHICK STANDARD; PRT; 1906 AA.  
 AC P11799; P19038;  
 DT 01-OCT-1989 (Rel. 12, Created)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Myosin light chain kinase, smooth muscle and non-muscle isozymes  
 DE (EC 2.7.1.117) (MCKK) [Contains: Telokin].  
 OS Gallus gallus (Chicken).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
 OC Gallus.  
 NC NCBI\_TaxID=9031;  
 RN [1]  
 RN SEQUENCE FROM N.A. (ISOFORM MLCK-210).  
 RX MEDLINE=96033976; PubMed=7589469;  
 RA Watterson D.M., Collinge M., Lukas T.J., van Eldik L.J., Birukov K.G.,  
 RA Stepanova O.V., Shitinsky V.P.,  
 RT "Multiple gene products are produced from a novel protein kinase  
 RT transcritption region.";  
 RL FEBS Lett. 373:217-220(1995).  
 RN [2]  
 RN SEQUENCE FROM N.A. (ISOFORM MLCK-108).  
 RX MEDLINE=90192792; PubMed=2315320;  
 RA Olson N.J., Pearson R.B., Needleman D.S., Hurwitz M.J., Kemp B.E.,  
 RA Means A.R.,  
 RT "Regulatory and structural motifs of chicken gizzard myosin light  
 RT chain kinase.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 87:2284-2288(1990).  
 RN [3]  
 RN SEQUENCE OF 649-1906 FROM N.A., AND PARTIAL SEQUENCE.  
 RC TISSUE=Fibroblast;  
 RX MEDLINE=90361738; PubMed=2202734;  
 RA Shoemaker M.O., Lau W., Shattuck R.L., Kwiatkowski A.P.,  
 RA Matrisian P.E., Guerra-Santos L., Wilson E., Lukas T.J.,  
 RA van Eldik L.J., Watterson D.M.,  
 RT "Use of DNA sequence and mutant analyses and antisense  
 RT oligodeoxynucleotides to examine the molecular basis of nonmuscle  
 RT myosin light chain kinase autoinhibition, calmodulin recognition, and  
 RT activity.";  
 RL J. Cell Biol. 111:1107-1125(1990).  
 RN [4]  
 RN SEQUENCE OF 1259-1906 FROM N.A.  
 RC TISSUE=Gizzard;  
 RX MEDLINE=9157587; PubMed=3030394;  
 RA Gueriere V. Jr., Russo M.A., Olson N.J., Putkey J.A., Means A.R.,  
 RT "Domain organization of chicken gizzard myosin light chain kinase  
 RT deduced from a cloned cDNA.";  
 RL Biochemistry 25:8372-8381(1986).  
 RN [5]  
 RN SEQUENCE OF 1750-1906 FROM N.A. (ISOFORM TELOKIN).  
 RC TISSUE=Gizzard;  
 RX MEDLINE=93073972; PubMed=1444462;  
 RA Yoshikai S., Ikebe M.,  
 RT "Molecular cloning of the chicken gizzard telokin gene and cDNA.";  
 RL Arch. Biochem. Biophys. 299:242-247(1992).  
 RN [6]  
 RN SEQUENCE OF 1750-1906 FROM N.A. (ISOFORM TELOKIN).  
 RX MEDLINE=92236611; PubMed=1373815;  
 RA Collinge M., Matrisian P.E., Zimmer W.E., Shattuck R.L., Lukas T.J.,  
 RA van Eldik L.J., Watterson D.M.,  
 RT "Structure and expression of a calcium-binding protein gene contained  
 RT within a calmodulin-regulated protein kinase gene.";  
 RL Mol. Cell. Biol. 12:2359-2371(1992).  
 CC -1- FUNCTION: Phosphorylates a specific serine in the N-terminus of a  
 CC myosin light chain, which leads to the formation calmodulin/MLCK  
 CC signal transduction complexes which allow selective transduction  
 CC of calcium signals.  
 CC -1- CATALYTIC ACTIVITY: ATP + [myosin light-chain] = ADP + [myosin  
 CC light-chain] phosphate.  
 CC -1- ALTERNATIVE PRODUCTS:

CC Event-Alternative Initiation; MLCK-210/Non-muscle.  
 CC Comment-At least 3 isoforms. MLCK-210/Non-muscle.  
 CC MLCK-108/Smooth-muscle and Telokin, are produced by alternative  
 CC Initiation;  
 CC -1- TISSUE SPECIFICITY: ISOFORM TELOKIN IS EXPRESSED IN GIZZARD,  
 CC HEART, LUNG, INTESTINE, AND SKELETAL MUSCLE ALTHOUGH THE LEVELS OF  
 CC THE EXPRESSION IN THE LATTER WERE MUCH LESS THAN THAT IN THE  
 CC GIZZARD.  
 CC -1- DOMAIN: TELOKIN BINDS CALMODULIN.  
 CC -1- SIMILARITY: Belongs to the Ser/Thr family of protein kinases.  
 CC -1- SIMILARITY: Contains 9 immunoglobulin-like C2-type domains.  
 CC -1- SIMILARITY: Contains 1 fibronectin type III domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 CC EMBL; X52876; CAA37056.1; -  
 CC EMBL; X52876; CAA37057.1; -  
 CC EMBL; X52876; CAA37058.1; -  
 CC EMBL; M31048; AAA49069.1; -  
 CC EMBL; M14953; AAA69964.1; -  
 CC EMBL; M96855; AAA49083.1; -  
 CC EMBL; M88283; AAA48647.1; -  
 CC EMBL; M88284; AAB53768.1; -  
 CC PIR; S68235; S68235.  
 CC PDB; 1YRK; 31-AUG-94.  
 CC PDB; 1YRK; 27-APR-99.  
 CC InterPro; IPR008957; FN\_III-like.  
 CC InterPro; IPR003961; FN\_III.  
 CC InterPro; IPR007110; IG-like.  
 CC InterPro; IPR003598; IG\_C2.  
 CC InterPro; IPR000719; Prot\_kinase.  
 CC InterPro; IPR002290; Ser\_thr\_pkin\_AS.  
 CC InterPro; IPR000441; Fn3; 1.  
 CC Pfam; PF00047; Ig; 9.  
 CC Pfam; PF00069; pkinase; 1.  
 CC ProDom; PD000001; Prot\_kinase; 1.  
 CC SMART; SM00060; FN3; 1.  
 CC SMART; SM00408; IGC2; 8.  
 CC SMART; SM00220; S\_TKC; 1.  
 CC PROSITE; PS50835; IG\_LIKE; 9.  
 CC PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 CC PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 CC PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 CC Transferrase; Serine/threonine-protein kinase; Calmodulin-binding;  
 CC ATP-binding; Phosphorylation; Immunoglobulin domain; Repeat;  
 CC Alternative Initiation; 3D-structure.  
 CC CHAIN 1 1906  
 CC FT MYOSIN LIGHT CHAIN KINASE, ISOFORM MLCK-  
 CC FT 210.  
 CC CHAIN 935 1906  
 CC FT MYOSIN LIGHT CHAIN KINASE, ISOFORM MLCK-  
 CC FT 108.  
 CC CHAIN 1750 1906  
 CC FT MYOSIN LIGHT CHAIN KINASE, ISOFORM  
 CC FT TELOKIN.  
 CC FT INIT MET 935 935  
 CC FT FOR ISOFORM MLCK-108.  
 CC FT INIT MET 1750 1750  
 CC FT FOR ISOFORM TELOKIN.  
 CC FT DOMAIN 156 117  
 CC FT IG-LIKE C2-TYPE 1.  
 CC FT DOMAIN 156 244  
 CC FT IG-LIKE C2-TYPE 2.  
 CC FT DOMAIN 429 517  
 CC FT IG-LIKE C2-TYPE 3.  
 CC FT DOMAIN 521 613  
 CC FT IG-LIKE C2-TYPE 4.  
 CC FT DOMAIN 637 725  
 CC FT IG-LIKE C2-TYPE 5.  
 CC FT DOMAIN 735 830  
 CC FT IG-LIKE C2-TYPE 6.  
 CC FT DOMAIN 1084 1172  
 CC FT IG-LIKE C2-TYPE 7.  
 CC FT DOMAIN 1225 1313  
 CC FT IG-LIKE C2-TYPE 8.  
 CC FT DOMAIN 1330 1400  
 CC FT FIBRONECTIN TYPE-III.  
 CC FT DOMAIN 1459 1885  
 CC FT IG-LIKE C2-TYPE 9.  
 CC FT ATP (BY SIMILARITY).  
 CC FT NP\_BIND 1459 1467



DR Pfam; PF00047; Ig; 1.  
 DR SMART; SM00407; IgC1; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 1.  
 DR PROSITE; PSS0290; IG\_MHC; 1.  
 KW Immunoglobulin domain; Immunoglobulin C region.  
 FT NON TER 1 1  
 FT DOMAIN 6 99  
 FT DISULFID 27 85  
 FT DISULFID 103 103  
 SQ SEQUENCE 104 AA; 11254 MW; CE4B67868886203 CRC64;

Query Match  
 Best Local Similarity 4.7%; Score 161; DB 1; Length 104;  
 Matches 35; Conservative 15; Mismatches 48; Indels 4; Gaps 2;

QY 521 QPRE-POVYTLPPSRDELTKNOVSLTCLVKGYPSPDIAVESNGOPENYKTPPVLD 579  
 1 QPKSTPLTVFPSSSELEKKNKATLVCLISNPSGVTAMKANGPITQGVDTSNPTKE 60

QY 580 DGSFELYSLTVDKSRWQGNVSCSYMHALNHHYKSL 621  
 61 GNRFMASFLHLSQDWRSHNSFTQVTHE--GDVTKSL 99

Db 61 GNRFMASFLHLSQDWRSHNSFTQVTHE--GDVTKSL 99

RESULT 72  
 HB2D\_PIG STANDARD; PRT; 258 AA.  
 ID HB2D\_PIG  
 AC P15963;  
 DT 01-APR-1990 (Rel. 14, Created)  
 DT 01-APR-1990 (Rel. 14, Last sequence update)  
 DT 15-JUL-1999 (Rel. 38, Last annotation update)  
 DE SLA class II histocompatibility antigen, DQ haplotype D beta chain precursor.  
 OS Sus scrofa (Pig).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.  
 NX NCBI\_TaxID=9623;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90361905; PubMed=2391424;  
 RA Gustafsson K., Leguenn C., Hirsch F., Germana S., Pratt K.,  
 Sachs D.H.;  
 RT "Class II genes of miniature swine. IV. Characterization and  
 expression of two allelic class II DQB cDNA clones.";  
 RL J. Immunol. 145:1946-1951(1990).  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by, and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M31498; AAA31085.1; -  
 DR HSSP; P13760; 2SER.  
 DR InterPro; IPR007110; Ig-like.  
 DR InterPro; IPR003597; Ig\_c1.  
 DR InterPro; IPR003066; Ig\_Mc.  
 DR InterPro; IPR000353; MHC\_II\_beta.  
 DR Pfam; PF00047; Ig; 1.  
 DR Pfam; PF00969; MHC\_II\_beta; 1.  
 DR ProDom; PD000328; MHC\_II\_beta; 1.  
 DR SMART; SM00407; IgC1; 1.  
 DR PROSITE; PSS0835; IG\_LIKE; 1.  
 DR PROSITE; PSS0290; IG\_MHC; 1.  
 KW MHC II; Transmembrane; Glycoprotein; Signal.  
 FT SIGNAL 1 31  
 FT CHAIN 32 258  
 FT DOMAIN 32 123  
 FT DOMAIN 124 217  
 FT DOMAIN 218 227  
 SLA CLASS II HISTOCOMPATIBILITY ANTIGEN,  
 DQ HAPLOTYPE D BETA CHAIN.  
 EXTRACELLULAR BETA-1.  
 EXTRACELLULAR BETA-2.  
 CONNECTING PEPTIDE.

FT TRANSMEM 228 248  
 FT DOMAIN 249 258  
 FT DISULFID 44 108  
 FT DISULFID 146 202  
 FT CARBOHYD 48 48  
 SQ SEQUENCE 258 AA; 29262 MW; E3AC75110AED47C3 CRC64;

Query Match  
 Best Local Similarity 4.7%; Score 161; DB 1; Length 258;  
 Matches 46; Conservative 28; Mismatches 80; Indels 8; Gaps 4;

QY 456 WYDGVENVNAKTKREEQNSTYRVSVTLVHQPMLNGEKYCKCVSKALPAPIEKT 515  
 54 WSDVDR-IYNQGEFLRPDSMDGEYRAVPLGRPADYLNQKRALEQKRAELDTVCNHY 112

QY 516 SKAG--QPREPOVYTLPPSRDELTKNOVSLTCLVKGYPSPDIAVESNGOPEN-NYK 571  
 113 QIEEGTLQRRVQPTTISPSKAEALNHHNLLVCVATDYPQVQVFRNGQETAGVV 172

QY 572 TTPPVLDSDGFYLSKLTVDKSRWQGNVSCSYMHALNHHYKSL 613  
 173 STPLIRNGD--WTYQVLMELMLQGDVYTCRVEHSSLQ 211

Db 173 STPLIRNGD--WTYQVLMELMLQGDVYTCRVEHSSLQ 211

RESULT 73  
 VCAM1\_HUMAN STANDARD; PRT; 739 AA.  
 ID VCAM1\_HUMAN  
 AC P19320;  
 DT 01-NOV-1990 (Rel. 16, Created)  
 DT 01-NOV-1990 (Rel. 16, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Vascular cell adhesion protein 1 precursor (V-CAM 1) (CD106 antigen) (INCM-100).  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 NX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Umbilical vein;  
 RX MEDLINE=91016951; PubMed=1699207;  
 RA Polte T., Newman W., Gopal T.V.;  
 RT "Full length vascular cell adhesion molecule 1 (VCAM-1)."  
 RL Nucleic Acids Res. 18:5901-5901(1990).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=90090619; PubMed=2688898;  
 RA Osborn L., Hession C., Tizard R., Vassallo C., Luhowskyj S.,  
 RA Chi-Rosso G., Lobb R.;  
 RT "Direct expression cloning of vascular cell adhesion molecule 1, a  
 RT cytokine-induced endothelial protein that binds to lymphocytes.";  
 RL Cell 59:1203-1211(1989).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91352090; PubMed=1715583;  
 RA Cybulsky M.I., Fries J.W.U., Williams A.J., Sulten P., Eddy R.,  
 RA Byers W., Shows T., Gimbrone M.A. Jr., Collins T.;  
 RT "Gene structure, chromosomal location, and basis for alternative mRNA  
 RT splicing of the human VCAM1 gene.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 88:7859-7863(1991).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=91201302; PubMed=1707873;  
 RA Hession C., Tizard R., Vassallo C., Schiffer S.B., Goff D., Moy P.,  
 RA Chi-Rosso G., Luhowskyj S., Lobb R., Osborn L.;  
 RT "Cloning of an alternate form of vascular cell adhesion molecule-1  
 RT (VCAM1)."  
 RL J. Biol. Chem. 266:6682-6685(1991).  
 RN [5]  
 RP SEQUENCE FROM N.A., AND VARIANTS PHE-318; ALA-384; ALA-413 AND  
 RP LEU-716.  
 RA Rieder M.J., Amel T.Z., Carrington D.P., Ozuna M., Kuldanek S.A.,

RA Kumar N.R., Toth E.J., Yi O., Nickerson D.A.;  
 RL Submitted (Aug-2002) to the EMBL/GenBank/DBJ databases.  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM LONG).  
 RC TISSUE:Retinal pigment epithelium;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shennan C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uediri T.B., Toshiyuki S., Carrinci P., Prange C.,  
 RA Rana S.S., Lequellano N.A., Peters G.J., Adamson R.A., Mullaby S.J.,  
 RA Bork S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunatratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,  
 RA Vallatton D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smallus D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).  
 RN [7]  
 RP X-RAY CRYSTALLOGRAPHY (1.8 ANGSTROMS) OF 25-226.  
 RX MEDLINE=95147978; PubMed=7531291;  
 RA Jones E.Y., Haxios K., Bottemann M.J., Robinson R.C., Driscoll P.C.,  
 RA Edwards R.M., Clements J.M., Dudgeon T.J., Stuart D.I.;  
 RT "Crystal structure of an integrin-binding fragment of vascular cell  
 RT adhesion molecule-1 at 1.8-A resolution."  
 RL Nature 373:539-544 (1995).  
 RN [8]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.  
 RX MEDLINE=95296382; PubMed=7539925;  
 RA Wang J.-H., Pepinsky R.B., Stehle T., Liu J.-H., Karpusas M.,  
 RA Browning B., Osborn L.;  
 RT "The crystal structure of an N-terminal two-domain fragment of  
 RT vascular cell adhesion molecule 1 (VCAM-1): a cyclic peptide based on  
 RT the domain 1 C-D loop can inhibit VCAM-1-alpha 4 integrin  
 RT interaction."  
 RL Proc. Natl. Acad. Sci. U.S.A. 92:5714-5718 (1995).  
 RN [9]  
 RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS) OF 25-220.  
 RA Wang J.-H., Stehle T., Pepinsky R.B., Liu J.-H., Karpusas M.,  
 RA Osborn L.;  
 RT "Structure of a functional fragment of VCAM-1 refined at 1.9-A  
 RT resolution."  
 RL Acta Crystallogr. D 52:369-379 (1996).  
 CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION  
 CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1  
 CC INTEGRIN VL44 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL  
 CC TRANSDUCTION. THE VCAM1/VL44 INTERACTION MAY PLAY A  
 CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE  
 CC EMIGRATION TO SITES OF INFLAMMATION.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing: Named isoforms=2;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=Long;  
 CC IsoId=PI9320-1; Sequence=Displayed;  
 CC Name=Short;  
 CC IsoId=PI9320-2; Sequence=VSP\_002580;  
 CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as  
 CC well as on macrophage-like and dendritic cell types in both normal  
 CC and inflamed tissue.  
 CC -1- INDUCTION: By cytokines (e.g. IL-1, TNF-alpha).  
 CC -1- PTM: Sialoglycoprotein.  
 CC -1- DISEASE: May play an important role in the genesis of  
 CC atherosclerosis and rheumatoid arthritis.  
 CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.

CC -1- DATABASE: NAME=PROV; NOTE=CD guide CD106 entry;  
 CC WWW="http://www.ncbi.nlm.nih.gov/prov/cd/cd106.htm".  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.ebi.ch/announce/>  
 CC or send an email to [license@ebi-sib.ch](mailto:license@ebi-sib.ch)).  
 CC -----  
 CC EMBL; X53051; CAA37218.1; -;  
 CC EMBL; M30257; AAA51917.1; ALT\_TERM.  
 CC EMBL; M73255; AAA61270.1; -;  
 CC EMBL; M60335; AAA61269.1; -;  
 CC EMBL; AF536818; AA096190.1; -;  
 CC EMBL; BC017276; AA017276.1; -;  
 CC PIR; A41288; A41288.  
 CC PIR; B41288; B41288.  
 CC PDB; 1VCA; 15-SRP-95.  
 CC PDB; 1VSC; 20-TUN-96.  
 CC PDB; 1I3J; 07-NOV-01.  
 CC Genew; HGNC:12663; VCAM1.  
 CC MIM; 192225; -;  
 CC InterPro; IPR003987; ICAM VCAM-1.  
 CC InterPro; IPR007110; Ig-like.  
 CC InterPro; IPR003598; Ig-C2.  
 CC InterPro; IPR003989; VCAM-1.  
 CC Pfam; PF00047; Ig; 6.  
 CC PRINTS; PR01472; ICAMVCAM1.  
 CC PRINTS; PR01474; VCAM1.  
 CC SMART; SM00408; IGC2; 3.  
 CC PROSITE; PS50835; IG\_LIKE; 5.  
 CC Immunoglobulin domain; Glycoprotein; Cell adhesion; Transmembrane;  
 CC Repeat; Signal; Alternative splicing; Polymorphism; 3D-structure.  
 CC CHAIN 1 24  
 CC SIGNAL 25 739  
 CC TRANSMEM 25 698  
 CC DOMAIN 699 720  
 CC DOMAIN 721 739  
 CC DOMAIN 25 105  
 CC DOMAIN 109 212  
 CC DOMAIN 223 309  
 CC DOMAIN 312 399  
 CC DOMAIN 408 506  
 CC DOMAIN 511 595  
 CC DOMAIN 600 684  
 CC DISULFID 47 95  
 CC DISULFID 52 99  
 CC DISULFID 137 195  
 CC CARBOHYD 273 273  
 CC CARBOHYD 365 365  
 CC CARBOHYD 417 417  
 CC CARBOHYD 463 463  
 CC CARBOHYD 531 531  
 CC CARBOHYD 561 561  
 CC CARBOHYD 610 610  
 CC VARSPLIC 310 402  
 CC VARIANT 318 318  
 CC VARIANT 384 384  
 CC VARIANT 413 413  
 CC VARIANT 716 716  
 CC STRAND 26 30  
 CC STRAND 34 38  
 CC TURN 39 40  
 CC STRAND 43 50  
 CC STRAND 56 61  
 CC TURN 62 63  
 CC STRAND 70 74  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC N-LINKED (GLCNAC. . .) (POTENTIAL).  
 CC Missing (in isoform Short).  
 CC S -> F.  
 CC /FTID=VAR\_013309.  
 CC T -> A.  
 CC /FTID=VAR\_014310.  
 CC G -> A.  
 CC /FTID=VAR\_014311.  
 CC I -> L.  
 CC /FTID=VAR\_014312.

FT TURN 75 76  
 FT STRAND 77 82  
 FT HELIX 87 88  
 FT STRAND 92 99  
 FT TURN 100 101  
 FT STRAND 102 114  
 FT STRAND 120 123

Query Match 4.78; Score 160.5; DB 1; Length 739;  
 Best Local Similarity 21.58; Pred. No. 0.0081;  
 Matches 118; Conservative 91; Mismatches 217; Indels 123; Gaps 27;

8 RHLLVQLALPLPAAT--QGNKVVLGGKDPVELTCTAGSQKSIQPHMKSNQIKILGNQ 65  
 211 KQAVKEIQVYISPKNTYISVSPSTKIQEGGSVTMTCSSEGLPAPELFW----- 258  
 66 GSFLLTKGPKSLNDRASRLMDQGNFLLIKKLIKEDSDTYICE---VEDQKEVOLT 121  
 259 -----SKKLDNGNLQHL---SGNATLTLIMRMEDSGIYCEGVNLIKRRKEVELI 307  
 122 V----FGLTANSDTHLQ--GQSILTLLESPGSSPSVQCRSPRGKNIQ-----GKT 168  
 308 VQKPFVEISPGRIAAQIGDSVMTLCSVWGCESPSPFSWRTQIDSPISGKVRSEGTNST 367  
 169 LSVSQLELDQSGMTCTVLQNKVKPKIDIVVLAFOKASSIYK---KEGEVRS--- 222  
 368 LTLSPVSENEHSTLCTVTCGKKLEKGIQVELYSPRDEPIEMSGGLVNGSSVTWSCV 427  
 223 ---FPL-AFTVEKLTSSEGLMWAERASSKSWITF---DLNKEVSVKRVTDPELQ- 273  
 428 PSVYPLDRLEIELLKGETIL-----ENIEFEDTDMSLENKSLKEMTFPIETED 476  
 274 MGKLL---PLHLTPALQVYAGSNLTLALBAKTKLHQBENLVVMATQIQOK---N 325  
 477 TGRALVQCAKLTIDMEFEKROKOS---TOTLVYVNAF--RDTVLVSPSIIIEGSSVN 531  
 326 LTCEWVGPTSPKMLSLKENKEKVKRKPWVNL---PEAGMOCQ-LSDSGOVLLE 381  
 532 MTCISQGFPAKPKLMSKQPLNGEIQPLSENAATLTLISTKMGDSGVYLCSEGINQGRKKE 551  
 382 SNIKVLEPTWSTPVPAPCPAPKPSCKTHTCPELLGGPSVFLFPP---KPKDTLMIISRT-- 435  
 592 VEILIQVT-----PKDKILT-----AFPSVSEVGGTVIISCTCGN 627  
 436 -PEVTCVVDVSHEDPEVKRMVYDVY-EVHNAKTRP---EQNSTYRVVSLTVLH 489  
 628 VPE-TWIIKKKAKETGDTVLK-SIDGAYTIRKQIKDAGVYECESKRVKSQLSRLTDV 685  
 490 QDMUNGKEY 498  
 686 QGRBNKKQY 694

RESULT 74  
 ID\_LAC\_HUMAN STANDARD; PRT; 105 AA.  
 AC P01842; P80423;  
 DT 21-JUL-1986 (Rel. 01, Created)  
 DT 21-JUL-1986 (Rel. 01, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE IG lambda chain C regions.  
 GN IGLC1 AND IGLC2 AND IGLC3.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Cranialia; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
 OX NCBI\_Taxid=9606;  
 RN [1]  
 RP SEQUENCE (BENCE-JONES PROTEIN SH).  
 RX MEDLINE=7016723; PubMed=4909564;  
 RA Titani K., Wikler M., Shinoda T., Putnam F.W.;  
 RT "The amino acid sequence of a lambda type Bence-Jones protein. 3. The  
 RT complete amino acid sequence and the location of the disulfide  
 RT bridges.";

RL J. Biol. Chem. 245:2171-2176 (1970).  
 [2]  
 RP SEQUENCE (BENCE-JONES PROTEIN X), AND DISULFIDE BONDS.  
 RX MEDLINE=69089380; PubMed=4883841;  
 RA Miletich C., Clegg J.B., Jarvis J.M.;  
 RT "Immunoglobulin lambda chains. The complete amino acid sequence of a  
 RT Bence-Jones protein";  
 RL Biochem. J. 110:631-652 (1966).  
 [3]  
 RP SEQUENCE (BENCE-JONES PROTEIN NIG-64).  
 RX MEDLINE=83186114; PubMed=6404900;  
 RA Kametani F., Takayasu T., Suzuki S., Shinoda T., Okuyama T.,  
 RA Shimizu A.;  
 RT "Comparative studies on the structure of the light chains of human  
 RT immunoglobulins. IV. Assignment of a subgroup";  
 RL J. Biochem. 93:421-429 (1983).  
 [4]  
 RP SEQUENCE (BENCE-JONES PROTEIN KERN).  
 RX MEDLINE=71150336; PubMed=5549568;  
 RA Ponetnig H., Hesse M., Hilschmann N.;  
 RT "Structural rule of antibodies. Primary structure of a monoclonal  
 RT immunoglobulin-L-chain of the lambda type, subgroup IV (Bence-Jones-  
 RT protein Kern). V. The complete amino acid sequence and its genetic  
 RT interpretation";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 352:247-266 (1971).  
 [5]  
 RP PARTIAL SEQUENCE (MYELOMA PROTEIN NEMM).  
 RX MEDLINE=74109253; PubMed=4814727;  
 RA Chen B.L., Poljak R.J.;  
 RT "Amino acid sequence of the (lambda) light chain of a human myeloma  
 RT immunoglobulin (IgG New).";  
 RL Biochemistry 13:1295-1302 (1974).  
 [6]  
 RP SEQUENCE (DOT).  
 RX MEDLINE=95355298; PubMed=7737190;  
 RA Stoppin M., Bellotti V., Negri A., Merlini G., Garver F., Ferri G.;  
 RT "Characterization of the two unique human anti-flavin monoclonal  
 RT immunoglobulins";  
 RL Eur. J. Biochem. 228:886-893 (1995).  
 [7]  
 RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF NEMM.  
 RX MEDLINE=75046825; PubMed=4215080;  
 RA Poljak R.J., Amzel L.M., Avey H.P., Chen B.L., Phizackerley R.P.,  
 RA Saul F.;  
 RT "The three-dimensional structure of the fab' fragment of a human  
 RT myeloma immunoglobulin at 2.0-A resolution";  
 RL Proc. Natl. Acad. Sci. U.S.A. 71:3440-3444 (1974).  
 [8]  
 RP PARTIAL SEQUENCE (BENCE-JONES PROTEIN MCG).  
 RX MEDLINE=75013804; PubMed=4415202;  
 RA Fect J.W., Deutsch H.F.;  
 RT "Primary structure of the Mcg lambda chain";  
 RL Biochemistry 13:4102-4114 (1974).  
 [9]  
 RP X-RAY CRYSTALLOGRAPHY (2.3 ANGSTROMS) OF MCG.  
 RX Edmundson A.B., Ely K.R., Abola E.E., Schiffer M.,  
 RA Panagiotopoulos N.;  
 RT "Rotational allomerism and divergent evolution of domains in  
 RT immunoglobulin light chains";  
 RL Biochemistry 14:3953-3961 (1975).  
 [10]  
 RP X-RAY CRYSTALLOGRAPHY OF MCG.  
 RX MEDLINE=90133913; PubMed=2515285;  
 RA Ely K.R., Herion J.N., Harter M.,  
 RT "Three-dimensional structure of a light chain dimer crystallized in  
 RT water. Conformational flexibility of a molecule in two crystal  
 RT forms";  
 RL J. Mol. Biol. 210:601-615 (1989).  
 [11]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=82080680; PubMed=6273747;  
 RA Hietar P.A., Hollis G.F., Kormeyer S.J., Waldmann T.A., Leder P.;  
 RT "Clustered arrangement of immunoglobulin lambda constant region genes



```

RT in man."
RL Nature 294:536-540 (1981).
CC -I- MISCELLANEOUS: The sequence shown is the Kern-/Oz-/MCG- chain
CC found in proteins SH, X, and NIG-64. The Kern protein has the
CC Kern+ marker, the NEMW protein has the Oz+ marker, the MCG protein
CC has the Kern+ marker, and the MCG+ marker.
CC -I- MISCELLANEOUS: Six tandem lambda-type genes were identified and
CC the 5' were sequenced. These correspond to the MCG sequence
CC (lambda-1), the Kern-/Oz- sequence (lambda-2) and the Kern-/Oz+
CC sequence (lambda-3).
CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements/
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; J00253; AAA59107.1; -.
DR EMBL; L38562; AAB36581.1; ALT_INT.
DR EMBL; X51754; CAB38569.1; ALT_INT.
DR EMBL; X51755; CAB36049.1; -.
DR EMBL; X51755; CAB36051.1; -.
DR PIR; A92057; L2HU.
DR PDB; 2MCG; 15-JUL-92.
DR PDB; 7FAB; 31-JAN-94.
DR PDB; 1AOK; 04-FEB-99.
DR PDB; 1LIL; 15-MAY-97.
DR Genew; HGNC:5855; IGLC1.
DR Genew; HGNC:5856; IGLC2.
DR Genew; HGNC:5857; IGLC3.
DR MIM; 147220; -.
DR GO; GO:0003823; F:antigen binding; NAS.
DR GO; GO:0006955; P:immune response; NAS.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00407; IGL1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; Bence-Jones protein;
KW 3D-structure.
FT 1 100 IG-LIKE.
FT 6 86 INTERCHAIN (WITH HEAVY CHAIN).
FT 27 104 DISULFID
FT 104 104
FT 5 5 A -> N (IN MCG+ MARKER).
FT 7 7 /FTID=VAR_003698.
FT 7 7 S -> T (IN MCG+ MARKER).
FT 45 45 /FTID=VAR_003899.
FT 45 45 S -> G (IN KERN+ MARKER).
FT 56 56 /FTID=VAR_003900.
FT 56 56 T -> K (IN MCG+ MARKER).
FT 82 82 /FTID=VAR_003901.
FT 82 82 R -> K (IN OZ+ MARKER).
FT 82 82 /FTID=VAR_003902.
FT STRAND 8 11
FT HELIX 15 19
FT TURN 20 21
FT STRAND 24 32
FT STRAND 38 43
FT TURN 44 45
FT STRAND 52 59
FT STRAND 61 62
FT TURN 65 72
FT HELIX 75 80
FT STRAND 84 89
FT STRAND 94 99
SO SEQUENCE 105 AA; 11236 MW; DCD9C7C201C13CC2 CRC64;

```

```

Query Match 4.7%; Score 159; DB 1; Length 105;
Best Local Similarity 34.1%; Pred. No. 0.00079;
Matches 31; Conservative 18; Mismatches 40; Indels 2; Gaps 2;

Qy 521 QPR-EPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-ENNYKTTTPVL 578
Db 1 QPKAPAVTLTFPSESELGANKATLVCLISDFYGAVTVMKADSPVKAGVETTTPSKQ 60
Qy 579 SDGSFPLYSKLTVDKSRMOCGNVFCGVME 609
Db 61 SNNRYAASSTYALSTPEQWKSHRSYSCVTHE 91

RESULT 75
ID LAC_PIG STANDARD; PRT; 105 AA.
AC P01846;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda chain C region.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE.
RA MEDLINE=7800254; PubMed=409425;
RA Novotny J., Franek F., Maggolis M.N., Haber E.;
RT "Amino acid sequence of normal (microheterogeneous) porcine
RL Biochemistry 16:3765-3772(1977).
CC -I- MISCELLANEOUS: This chain was obtained from a mixture of normal
CC immunoglobulins.
CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.
DR PIR; A02129; LIPG.
DR HSSP; P01842; 2MCG.
DR InterPro; IPR007110; IG-like.
DR InterPro; IPR003597; IG-cl.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig; 1.
DR SMART; SM00407; IGL1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT 1 100 IG-LIKE.
FT 6 86 INTERCHAIN (WITH HEAVY CHAIN).
FT 27 104 DISULFID
FT 104 104
FT 5 5 A -> N (IN MCG+ MARKER).
FT 7 7 /FTID=VAR_003698.
FT 7 7 S -> T (IN MCG+ MARKER).
FT 45 45 /FTID=VAR_003899.
FT 45 45 S -> G (IN KERN+ MARKER).
FT 56 56 /FTID=VAR_003900.
FT 56 56 T -> K (IN MCG+ MARKER).
FT 82 82 /FTID=VAR_003901.
FT 82 82 R -> K (IN OZ+ MARKER).
FT 82 82 /FTID=VAR_003902.
FT STRAND 8 11
FT HELIX 15 19
FT TURN 20 21
FT STRAND 24 32
FT STRAND 38 43
FT TURN 44 45
FT STRAND 52 59
FT STRAND 61 62
FT TURN 65 72
FT HELIX 75 80
FT STRAND 84 89
FT STRAND 94 99
SO SEQUENCE 105 AA; 11003 MW; 3817AABD74C396 CRC64;

Query Match 4.6%; Score 156; DB 1; Length 105;
Best Local Similarity 35.2%; Pred. No. 0.0012;
Matches 32; Conservative 15; Mismatches 42; Indels 2; Gaps 2;

Qy 521 QPR-EPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQP-ENNYKTTTPVL 578
Db 1 QPKAPAVTLTFPSESELGANKATLVCLISDFYGAVTVMKADSPVKAGVETTTPSKQ 60
Qy 579 SDGSFPLYSKLTVDKSRMOCGNVFCGVME 609
Db 61 SNNRYAASSTYALSTPEQWKSHRSYSCVTHE 91

RESULT 76
ID LACS_MUSSP STANDARD; PRT; 105 AA.
AC P20765;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda-5 chain C region.

```

```

OS Mus spretus (Western wild mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10096;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=88196070; PubMed=3129289;
RA Mami F., Cazeneuve P.A., Kind T.J.;
RL "Conservation of the immunoglobulin C lambda 5 gene in the Mus gene.";
EMBO J. 7:117-122(1988).
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M35582; AAA39152.1; -.
CC HSSP; P01842; 2MCG.
CC InterPro; IPR007110; IG_1like.
CC InterPro; IPR003597; IG_c1.
CC InterPro; IPR003006; IG_MHC.
CC Pfam; PF00047; Ig_1.
CC SMART; SM00407; IgC1; 1.
CC PROSITE; PS50835; IG_LIKE; 1.
CC PROSITE; PS00290; IG_MHC; 1.
CC Immunoglobulin domain; Immunoglobulin C region.
CC NON_TER 1 1
CC DOMAIN 6 100 IG-LIKE.
CC DISULFID 27 86
CC DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
CC SEQUENCE 105 AA; 11674 MW; AAB417DF68471A17 CRC64;

Query Match 4.5%; Score 153.5; DB 1; Length 105;
Best Local Similarity 34.0%; Pred. No. 0.0018; Indels 5; Gaps 3;
Matches 35; Conservative 20; Mismatches 43;

QY 521 QPR-EPQVYTLPPSRDELTKNOVSLTCLVKGFFYPSDIAVEMESNGQP-ENNYKTPPVLD 578
DB 1 QPRSDPLVTLPLPSIKNLQKANKTVLCLVSEFPGTLVDMKVDGVTVGTGVTTPSQ 60
QY 579 SDGSFFLYSKLVYDKSRWQGNVFCGVMEALHNHYTKSL 621
DB 61 TNNKYMVSSYLTLISDQMPHSHSYSCRYTHE---GNTVERKSVS 100

RESULT 77
LAC RABIT STANDARD; PRT; 105 AA.
ID LAC RABIT
AC P01847;
DT 21-JUL-1986 (Rel. 01, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig lambda chain C region.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE.
RX MEDLINE=82091105; PubMed=6797414;
RA Garcia I., Jaton J.-C.;
RT "The primary structure of the constant region of Basilea-rabbit
RT immunoglobulin lambda-chains.";
RL Biochem. J. 197;177-183(1981).
CC -1- MISCELLANEOUS: This lambda chain expresses the c7 allotypic
CC specificity.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC PIR; A02130; L7RB.
CC HSSP; P01842; 7FAB.

```

```

DR InterPro; IPR007110; IG_1like.
DR InterPro; IPR003597; IG_c1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IgC1; 1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; FALSE NEG.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 6 100 IG-LIKE.
FT DISULFID 27 86
FT DISULFID 104 104 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 105 AA; 11484 MW; B427513272E863D CRC64;

Query Match 4.5%; Score 153; DB 1; Length 105;
Best Local Similarity 32.7%; Pred. No. 0.0019;
Matches 32; Conservative 17; Mismatches 45; Indels 4; Gaps 2;

QY 525 PQVYTLPPSRDELTKNOVSLTCLVKGFFYPSDIAVEMESNGQP-ENNYKTPPVLDGSP 583
DB 6 PQVYTLPPSRDELTKNKKATLVCLISDFRYTKVKKMKADGNSVTGQVDTPPSKSNKK 65
QY 584 FLYSKLVYDKSRWQGNVFCGVMEALHNHYTKSL 621
DB 66 AASSFLHLTANQWKSQSVTCQVTHE---GHTVERKSLA 100

RESULT 78
LAC RABIT STANDARD; PRT; 106 AA.
ID LAC RABIT
AC P01839;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Ig kappa-B4 chain C region.
GN K-BAS.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxId=9986;
RN [1]
RP SEQUENCE FROM N.A.
RA Heidmann O., Rougeon F.;
RT "Multiplicity of constant kappa light chain genes in the rabbit
RT genome: a b4b homozygous rabbit contains a kappa-bas gene";
RL EMO J. 2:437-441(1983).
CC -1- MISCELLANEOUS: In Basilea rabbits, the major type of light chain
CC is lambda. The kappa chain shown is a minor component. All other
CC rabbit B allotypes have Cys-64.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; V01241; CAA24558.1; -.
CC EMBL; V00885; -. NOT_ANNOTATED_CDS.
CC PIR; A02121; K4RBS.
CC HSSP; P01842; 7FAB.
CC InterPro; IPR007110; IG_1like.
CC InterPro; IPR003597; IG_c1.
CC InterPro; IPR003006; IG_MHC.
CC Pfam; PF00047; Ig_1.
CC SMART; SM00407; IgC1; 1.
CC PROSITE; PS50835; IG_LIKE; 1.
CC PROSITE; PS00290; IG_MHC; FALSE NEG.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1

```

FT DOMAIN 6 99 IG-LIKE.  
 FT DISULFID 27 87  
 FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).  
 SO SEQUENCE 106 AA; 11279 MW; AF9B928DA853849 CRC64;  
 Query Match 4.5%; Score 153; DB 1; Length 106;  
 Best Local Similarity 33.7%; Pred. No. 0.0019;  
 Matches 31; Conservative 19; Mismatches 36; Indels 6; Gaps 2;  
 QY 520 GQPEPEQYVTLTPSRDELTKNOVSLTCLVKGFYPSDIAVEKESNGOE---NNYKTPP 575  
 DB 1 GDPVAPSVLLFPSPKELTGTATVCAANKFYSIDITVWKVDGTTQSGSIENSKT--P 58  
 QY 576 VLDSGSFPLYSKLTVDKSRQGNVFSQSYV 607  
 DB 59 QSPEDNTYSLSTSLTSQAQNSHSVYTCFV 90  
 RESULT 79  
 ID CD22 HUMAN STANDARD; PRT: 847 AA.  
 AC P20273; O95699; O95701; O95702; Q01665; Q92872; Q92873;  
 AC Q9UQA1; Q9UQA8; Q9UQA9; Q9UQA9; Q9Y2A6;  
 DT 01-FEB-1991 (Rel. 17, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE B-cell receptor CD22 precursor (Ieu-14) (B-lymphocyte cell adhesion molecule) (Br-CM) (Siglec-2).  
 GN CD22.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RC SEQUENCE FROM N.A. (ISOFORM CD22-BETA), AND VARIANT HIS-639.  
 RC TISSUE=Testis;  
 RC MEDLINE=91086838; PubMed=1985119;  
 RA Wilson G.L., Fox C.H., Fauci A.S., Kehrl J.H.;  
 RT "CDNA cloning of the B cell membrane protein CD22: a mediator of B-B cell interactions.";  
 RT J. Exp. Med. 173:137-146(1991).  
 RN [2]  
 RC SEQUENCE FROM N.A. (ISOFORMS CD22-ALPHA AND CD22-BETA).  
 RC MEDLINE=93267103; PubMed=8496602;  
 RA Wilson G.L., Najfeld V., Kozlow E., Menniger J., Ward D.,  
 RA Kehrl J.H.;  
 RT "Genomic structure and chromosomal mapping of the human CD22 gene.";  
 RT J. Immunol. 150:5013-5024(1993).  
 RN [3]  
 RC SEQUENCE FROM N.A. (ISOFORMS CD22-ALPHA AND CD22-BETA).  
 RA Lamerstein J.E., McCreedy P., Adamson A.W., Burkhart-Schultz K.,  
 RA Garcia E., Kyle A., Ramirez M., Stillwagen S., Gaines J.,  
 RA Dangaran L., Bruce R., Quan G., Montgomery M., Ow D.,  
 RA Kobayashi A., Olsen A.O., Carrano A.V.;  
 RA Submitted (AUG-1996) to the EMBL/GenBank/DBJ databases.  
 RN [4]  
 RC SEQUENCE FROM N.A. (ISOFORM CD22-BETA).  
 RA MEDLINE=90231465; PubMed=1691828;  
 RA Stamenkovic I., Seed B.;  
 RT "The B-cell antigen CD22 mediates monocyte and erythrocyte adhesion.";  
 RT Nature 345:74-77(1990).  
 RN [5]  
 RC SEQUENCE OF 13-137, 139-239, 241-328 AND 418-502 FROM N.A., AND  
 RP VARIANTS THR-34; GLU-152; LYS-203; GLY-664; CYS-669 AND ASP-745.  
 RX MEDLINE=99180618; PubMed=10079291;  
 RA Hata Y., Teuchiya N., Matsushita M., Shiota M., Hagiwara K.,  
 RA Tokunaga K.;  
 RT "Identification of the gene variations in human CD22.";  
 RT Immunogenetics 49:280-286(1999).  
 RN [6]  
 RC SIALIC ACID BINDING.  
 RX MEDLINE=93216636; PubMed=8463235;  
 RA Powell L.D., Sgroi D., Sjöberg E.R., Stamenkovic I., Varzi A.;

RT "Natural ligands of the B cell adhesion molecule CD22 beta carry  
 RT N-linked oligosaccharides with alpha-2,6-linked sialic acids that are  
 RT required for recognition.";  
 RT J. Biol. Chem. 268:7019-7027(1993).  
 RN [7]  
 RC INTERACTION WITH PTPN6.  
 RX MEDLINE=95343349; PubMed=7618087;  
 RA Doody G.M., Juszczyk L.B., Delibrias C.C., Matthews R.J., Lin J.,  
 RA Thomas M.L., Fearon D.T.;  
 RT "A role in B cell activation for CD22 and the protein tyrosine  
 RT phosphatase SHP.";  
 RT Science 269:242-244(1995).  
 RN [8]  
 RC INTERACTION WITH LYN; SYK AND PIK3R1/PIK3R2.  
 RX MEDLINE=96257803; PubMed=8647200;  
 RA Tuscano J.W., Engel P., Tedder T.F., Agarwal A., Kehrl J.H.;  
 RT "Involvement of p72syk kinase, p53/56lyn kinase and phosphatidylyl  
 RT inositol-3 kinase in signal transduction via the human B lymphocyte  
 RT antigen CD22.";  
 RT Eur. J. Immunol. 26:1246-1252(1996).  
 RN [9]  
 RC INTERACTION WITH PTPN6; SYK AND PLCG1.  
 RX MEDLINE=96195207; PubMed=8627166;  
 RA Law C.L., Sidorenko S.P., Chandran K.A., Zhao Z., Shen S.H.,  
 RA Fischer E.H., Clark E.A.;  
 RT "CD22 associates with protein tyrosine phosphatase 1C, Syk, and  
 RT phospholipase C-gamma(1) upon B cell activation.";  
 RT J. Exp. Med. 183:547-560(1996).  
 RN [10]  
 RC REVIEW.  
 RX MEDLINE=97288746; PubMed=9143697;  
 RA Tedder T.F., Tuscano J., Sato S., Kehrl J.H.;  
 RT "CD22, a B lymphocyte-specific adhesion molecule that regulates  
 RT antigen receptor signaling.";  
 RT Annu. Rev. Immunol. 15:481-504(1997).  
 CC -1- FUNCTION: Mediates B-cell B-cell interactions. May be involved in  
 CC the localization of B-cells in lymphoid tissues. Binds sialylated  
 CC glycoproteins; one of which is CD45. Preferentially binds to  
 CC alpha2,6-linked sialic acid. The sialic acid recognition site can  
 CC be masked by cis interactions with sialic acids on the same cell  
 CC surface. Upon ligand induced tyrosine phosphorylation in the  
 CC immune response seems to be involved in regulation of B cell  
 CC antigen receptor signaling. Plays a role in positive regulation  
 CC through interaction with Src family tyrosine kinases and may also  
 CC act as an inhibitory receptor by recruiting cytoplasmic  
 CC phosphatases via their SH2 domains that block signal transduction  
 CC through dephosphorylation of signaling molecules.  
 CC -1- SUBUNIT: Predominantly monomer of isoform CD22-beta. Also found as  
 CC heterodimer of isoform CD22-beta and a shorter isoform. Interacts  
 CC with PTPN6/SHP-1, LYN, SYK, PIK3R1/PIK3R2 and PLCG1 upon  
 CC phosphorylation. Interacts with GRR2, INPP5D and SHC1 upon  
 CC INPP5D/SHIP, GRR2 and SHC1.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Comment=Additional isoforms seem to exist;  
 CC Name=CD22-beta;  
 CC IsoId=P20273-1; Sequence=Displayed;  
 CC Name=CD22-alpha;  
 CC IsoId=P20273-2; Sequence=VSP\_002531;  
 CC -1- TISSUE SPECIFICITY: B lymphocytes.  
 CC -1- DOMAIN: Contains 4 copies of a cytoplasmic motif that is referred  
 CC to as the immunoreceptor tyrosine-based inhibitor motif (ITIM).  
 CC This motif is involved in downmodulation of cellular responses.  
 CC The phosphorylated ITIM motif binds to the SH2 domain of  
 CC PTPN6/SHP-1.  
 CC -1- PTM: Phosphorylated both on threonine/serine and tyrosine  
 CC residues.  
 CC -1- PTM: Phosphorylated on tyrosine residues by LYN (By similarity).  
 CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily. SIGLEC  
 CC (sialic acid binding Ig-like lectin) family.  
 CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.

FT	CARBOHYD	67	67	N-LINKED (GLCNAC. . .) (POTENTIAL).
Query Match		4.5*	Score 152.5;	DB 1; Length 847;
Best Local Similarity		19.1*	Prod. No. 0.031;	
Matches 130;	Conservative	105;	Mismatches 229;	Indels 217; Gaps 36;
QY	LIINKLIEDSDPYICE-----		VEDOKEEVLVYGLTANSDTHL	133
DB	11 LVLEYLAFSSSKMVEFHEPETHLAMEGACWIMICTYRALDGDLESFILPFPNPEYKNKTSK			70
QY	134 LOGOSLTLTLE--SPSSSPSVCCRSBRKNIGGKTLISVQLDSDSGTCTCVLONOK			191
DB	71 FDGRLRYESTKDGKVPSEBOKRVQFLDKNKNC-----TLSTHPLHNDSSG--LGLRMES			123
QY	192 KVEEKIDIVL-----AFQKASIVYKKEGEV-----FSPPLAFVEKLT			233
DB	124 KTEMMERHILNVSERFPFHIDLPBI---QSGVEYTLTCLANFSCTGYFI-----			172
QY	234 GSGELVQWQERASSSKSWITFDLKNKEVSRYRTVDPDKLQMGKKLPLHLTLPOLPOYAG			293
DB	173 ---OLQWLTLGVEPMRQALVT---STSLTTSVFRSELSKFS-----PQMSH			212
QY	294 SGNL--TLAEAKTKGKL--HOEVLVYMRATQLO-----KVLTCGEWGPST			335
DB	213 HGKIVTQLODADGKFLSNDVTVLNKAHTPKLBIKVTPSDAIVREGDSVTMTCEV--SSSN			271
QY	336 PKLM-----LSLKLNKEAKVSKREKEVWVLNPEAGMQLLSD-----SGOVLLE			361
DB	272 PEYTVTSWMLDGLSKKQN--TFLTLNRE-----VTKQDSGKYCCQVNSDVPGRSEVFLQ			326
QY	382 -----SNIKVL-----PTWSTVPC-----PAPERSCKTH-----			408
DB	327 VQVAPEDSTVOIILHSPAVEGQVEFLCMSLNLPTNYTYHNGKEQWQTEERKVIH PKI			366
QY	409 -----TCPELIG--GGSVFL--FPPEPKDTLMISRTP-----EVTGVVDVSH			447
DB	387 LPMHAGTYSVCAENILITGQRGFGALDVQYPRKXTTVIQNMPIREGDTVLSCVNS			446
QY	448 EDPEV--KENNYVDGVEVNAKTKYPREBOYNSTYRVVSVLTVLHODVLNGKEYCKVSNK-			505
DB	447 SNPSVTREYEW-----KPHGAMSEPS-----LGVLTQIOWGMDN--TTIACAACNSW			490
QY	506 -ALPAPLEKTIISKAKQPREPOVYTLPSRDELTKQVSLTCLVKGFPSYSDIIVEMESG			564
DB	491 CSWASPPALANVOYA--PRDVRPRKTKPLSEIHSGNSVSLQCPSSSHPRKEVQPFMEKNG			547
QY	565 -----QENNNYKTPPYLDSGSEFFLYSKLTVDKSRWQOGNVSCSYMEBALHNYTQKS			619
DB	548 RLIGKESQLNFDSISF--EDAGSY-----SCVNNISIGQTASKAWTLLEVLAPRR			595
QY	620 L--SLSPGLDLDE-----TC 632			
DB	596 LRVSMSPGDQWMEGSKATLTTC 616			
RESULT 80				
DTC_HUMAN				
ID_DTC_HUMAN	STANDARD;	PRT;	383	AA.
AC	P01880;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	10-OCT-2003 (Rel. 42, Last annotation update)			
DE	IG delta chain C region.			
GN	IGHD.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid:9606;			
RN	[1]			
RP	SEQUENCE OF 1-162 (MYELOMA PROTEIN WAH).			
RX	MEDLINE=62082419; PubMed=6947220;			
RT	Putnam F.W., Takahashi N., Tetaert D., Debutre B., Lin L.-C.,			
RT	"Amino acid sequence of the first constant region domain and the			





OY 533 SDELTKNOVSLTCLVKGYPSDIWESNGCPENN---YKTPPV 576  
 DB 716 PENIDDESQVP-----DQPSLHV-----RPQTNCLIMSWTEPL 749  
 RESULT 83  
 CMLL\_HUMAN  
 ID CAML\_HUMAN STANDARD; PRT; 1257 AA.  
 AC P32004; Q8TA87;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-MAR-2004 (Rel. 43, Last annotation update)  
 DE Neutral cell adhesion molecule L1 precursor (N-CAM L1) (CD171 antigen).  
 GN L1CAM OR CAML1 OR MIC5.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92031698; PubMed=1932117;  
 RA Kobayashi M., Miura M., Asou H., Uyemura K.;  
 RT "Molecular cloning of cell adhesion molecule L1 from human nervous  
 tissue: a comparison of the primary sequences of L1 molecules of  
 RT different origin.";  
 RL Biochim. Biophys. Acta 1090:238-240(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=92120663; PubMed=1769655;  
 RA Hlavín M.L., Lemmon V.;  
 RT "Molecular structure and functional testing of human L1CAM: an  
 RT interspecies comparison.";  
 RL Genomics 11:416-423(1991).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=92329299; PubMed=1627459;  
 RA Reid R.A., Hemperly J.J.;  
 RT "Variants of human L1 cell adhesion molecule arise through alternate  
 RT splicing of RNA.";  
 RL J. Mol. Neurosci. 3:127-135(1992).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RA Rosenthal A., Coutelle O., Drescher B.;  
 RL Submitted (Apr-1994) to the EMBL/GenBank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=97432815; PubMed=9286695;  
 RA Brenner V., Nyakatura G., Rosenthal A., Platzter M.;  
 RT "Genomic organization of two novel genes on human Xq28: compact head  
 RT to head arrangement of IDH gamma and TRAP delta is conserved in rat  
 RT and mouse.";  
 RL Genomics 44:8-14(1997).  
 RN [6]  
 RP SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.  
 RX MEDLINE=98147998; PubMed=9479034;  
 RA Coutelle O., Nyakatura G., Taudien S., Elgar G., Brenner S.,  
 RA Platzter M., Drescher B., Jouet M., Kenwright S., Rosenthal A.;  
 RT "The neural cell adhesion molecule L1: genomic organization and  
 RT differential splicing is conserved between man and the pufferfish  
 RT Fugu.";  
 RL Gene 208:7-15(1998).  
 RN [7]  
 RP SEQUENCE OF 20-36.  
 RX MEDLINE=88298876; PubMed=3136168;  
 RA Wolff J.M., Frank R., Mujoo K., Spico R.C., Reisfeld R.A.,  
 RA Rachej F.G.;  
 RT "A human brain glycoprotein related to the mouse cell adhesion  
 RT molecule L1.";  
 RL J. Biol. Chem. 263:11943-11947(1988).  
 RN [8]  
 RP SEQUENCE OF 332-371 FROM N.A.

RX MEDLINE=90353957; PubMed=2387585;  
 RA Djabali M., Mattei M.-G., Nguyen C., Roux D., Demengeot J.,  
 RA Denizot F., Moos M., Schachner M., Goridis C., Jordan B.R.;  
 RT "The gene encoding L1, a neural adhesion molecule of the  
 RT immunoglobulin family, is located on the X chromosome in mouse and  
 RT man.";  
 RL Genomics 7:587-593(1990).  
 RN [9]  
 RP SEQUENCE OF 353-1176 FROM N.A.  
 RC TISSUE=Fetal brain;  
 RX MEDLINE=92020233; PubMed=1923824;  
 RA Rosenthal A., Mackinnon R.N., Jones D.S.C.;  
 RT "PCR walking from microdissection clone M54 identifies three exons  
 RT from the human gene for the neural cell adhesion molecule L1  
 RT (CAM-L1).";  
 RL Nucleic Acids Res. 19:5395-5401(1991).  
 RN [10]  
 RP SEQUENCE OF 809-1257 FROM N.A.  
 RC TISSUE=Pancreas;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Bluetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Haile F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Rana S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultky S.W.,  
 RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Gilmwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skaleka U., Smallie D.E.,  
 RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [11]  
 RP SEQUENCE OF 1030-1257 FROM N.A.  
 RX MEDLINE=91132183; PubMed=1993895;  
 RA Harper J.R., Prince J.T., Healy P.A., Stuart J.K., Nauman S.J.,  
 RA Stalcup W.B.;  
 RT "Isolation and sequence of partial cDNA clones of human L1: homology  
 RT of human and rodent L1 in the cytoplasmic region.";  
 RL J. Neurochem. 56:797-804(1991).  
 RN [12]  
 RP VARIANT HSAS TYR-264.  
 RX MEDLINE=94004956; PubMed=8401576;  
 RA Jouet M., Rosenthal A., Macfarlane J., Kenwright S., Donnai D.;  
 RT "A missense mutation confirms the L1 defect in X-linked hydrocephalus  
 RT (HSAS).";  
 RL Nat. Genet. 4:331-331(1993).  
 RN [13]  
 RP VARIANT HSAS/MASA LEU-1194.  
 RX MEDLINE=95187172; PubMed=7881431;  
 RA Friesen E., Schrander-Stumpel C., Vits L., Coucke P., van Camp G.,  
 RA Williams P.J.;  
 RT "X-linked hydrocephalus and MASA syndrome present in one family are  
 RT due to a single missense mutation in exon 28 of the L1CAM gene.";  
 RL Hum. Mol. Genet. 3:2255-2256(1994).  
 RN [14]  
 RP VARIANTS HSAS GLN-184 AND ARG-452, AND VARIANT MASA GLN-210.  
 RX MEDLINE=95004608; PubMed=7920659;  
 RA Jouet M., Rosenthal A., Armstrong G., Macfarlane J., Stevenson R.,  
 RA Paterson J., Metzger A., Ionescu V., Temple K., Kenwright S.;  
 RT "X-linked spastic paraplegia (SPG1), MASA syndrome and X-linked  
 RT hydrocephalus result from mutations in the L1 gene.";  
 RL Nat. Genet. 7:402-407(1994).  
 RN [15]  
 RP VARIANTS MASA GLN-210 AND ASN-598.

```

RX MEDLINE=95004609; PubMed=7920660;
RA Vits L., van Camp G., Coucke P., Franssen E., de Boule K.,
RA Reyniers E., Korn B., Poustra A., Wilson G., Schrander-Stumpel C.,
RA Wintter R.M., Schwartz C., Willems P.J.;
RT "MASA syndrome is due to mutations in the neural cell adhesion gene
RT LICAM."
RL Nat. Genet. 7:408-413(1994).
[16]
RN VARIANTS HSAS/MASA SER-9; SER-121; LYS-309; PHE-768; LEU-941 AND
RN CYS-1070.
RX MEDLINE=95282776; PubMed=7762552;
RA Jouet M., Moncla A., Paterson J., McKeown C., Fryer A., Carpenter N.,
RA Holmberg E., Medelius C., Kenwright S.;
RT "New domain of neural cell-adhesion molecule L1 implicated in
RT X-linked hydrocephalus and MASA syndrome."
RL Am. J. Hum. Genet. 56:1304-1314(1995).
[17]
RN VARIANTS HSAS/MASA GLN-184; GLN-210; TYR-264; ARG-452; ASN-598 AND
RN LEU-1194.
RX MEDLINE=96153146; PubMed=8556302;
RA Franssen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RT "CRASH syndrome: clinical spectrum of corpus callosum hypoplasia,
RT ectodactyria, adducted thumbs, spastic paraparesis and hydrocephalus
RT due to mutations in one single gene, L1."
RL Eur. J. Hum. Genet. 3:273-284(1995).
[18]
RN ERRATUM.
RA Franssen E., Lemmon V., van Camp G., Vits L., Coucke P., Willems P.J.;
RL Eur. J. Hum. Genet. 4:126-126(1996).
[19]
RN VARIANTS HSAS/MASA/SPG1 SER-179 AND ARG-370.
RX MEDLINE=96057511; PubMed=7562969;
RA Ritz J.C., Cuppens H., Legius E., Fryns J.-P., Glover T., Marynen P.,
RA Caessman J.-J.;
RT "Mutations in L1-CAM in two families with X linked complicated
RT spastic paraplegia, MASA syndrome, and HSAS."
RL J. Med. Genet. 32:549-552(1995).
[20]
RN VARIANTS HSAS CYS-194 AND LEU-240.
RX MEDLINE=97083370; PubMed=8929944;
RA Gu S.-M., Orth U., Veeke A., Enders H., Klunder K., Schlosser M.,
RA Engel W., Schwinger E., Gal A.;
RT "Five novel mutations in the LICAM gene in families with X linked
RT hydrocephalus."
RL J. Med. Genet. 33:103-106(1996).
[21]
RN VARIANTS HSAS GLN-184; 439-VAL--THR-443 DEL; CYS-784 AND
RN 936-LEU--LEU-948 DEL.
RX MEDLINE=97338664; PubMed=9195224;
RA Macfarlane J.R., Du J.-S., Peyys M.E., Ramsden S., Domai D.,
RA Charlton R., Garrett C., Tolve U., Yates J.R.W., Berry C., Goudie D.,
RA Moncla A., Lunt P., Hodgson S., Jouet M., Kenwright S.;
RT "Nine novel L1 CAM mutations in families with X-linked
RT hydrocephalus."
RL Hum. Mutat. 9:512-518(1997).
[22]
RN VARIANTS HSAS/MASA ASP-691; ARG-698 AND PRO-935.
RX MEDLINE=98180721; PubMed=9521424;
RA Du Y.-Z., Srivastava A.K., Schwartz C.E.;
RT "Multiple exon screening using restriction endonuclease
RT fingerprinting (REF): detection of six novel mutations in the L1 cell
RT adhesion molecule (LICAM) gene."

```

```

Db 469 IODERFF-----PYANGTIGIRDIQANDTGRVFCILANDONNVITM 509
Qy 122 VEGILANDSDTHLQ-----GSLTLTLLES--PQSSPSVQGRSPRGKNIQ----- 164
Db 510 A-NLVKQATQITQGPRTSTTEKGRVITFTQASDFPSIQPSITWKGCD-GRDQELGDS 567
Qy 165 -----GKTLISVSOGLQDOSTGTCTVQLQCKKVEFKIDIVLAF-----QKASSIVYKK 214
Db 566 KVFIEDGR-LVHSHLDYDQGNVSCVASTELDVESRAGLLVGSRGPRVLVSLHL 626
Qy 215 EGEQVEFSPFLA-----FTVEKLTSGSELMOAERASSSKSWITFDLKNKEVS 263
Db 627 TOSQVRVSWSPADENHAPIEKYDIEFEDKEMAPBKWYSLGKVPNGQSTLTLSPYVHYT 686
Qy 264 KVTQDPKIQMKKPLNHLTL--PQALPQ-----YAGSNTLLALBAKTKGLHGVNLV 316
Db 687 FVTALINKYGPSPVSEVTVPFAAPKPNVDVKBGENE-----TNVYI 733
Qy 317 MRATQLOKULTCGVGP-----TSPKMLSLKENKAKVSKREKPVVVLNPEAGMNO- 369
Db 734 -----TWKPLRMMDMNAPOVQYRQ-----WRPGTRGPQGE 765
Qy 370 CILSDSGOVLLESNIKVLPTWSTPVPCC-----PAPEKSC-----DKHTCPRELIG-- 415
Db 766 QIVSD--PFLVVSNTSTFVPEIKVQAVNSGKGPEPQVTIGSGEDYPQALPELEGIRI 823
Qy 416 -GPSVFLPPPKKDTLMIS--RTPEVTCVVVDVSHEDDEVFNMYVDGVEVHNNAKTKR 471
Db 824 LNSSAVLVKMRPVDLAQVNGHLRGYNVT-----TWRESQKHSRGH 867
Qy 472 EBOY-----NSTRVVSVLTVLHQMNGKVKCKVSNKALPAPIEKTISKAGOPREPQV 527
Db 868 KHVYVPANTTISVILSGLRPYSYHLEVOAFNGRSGRA-----SETFTPEPGVPHRPA 923
Qy 528 YTL-----PESRDELTKNOVSLTGLVKGFPYSDIAVENESGQBPENNYKTP 574
Db 924 LHLGCSNTSLLRMQP-----LSHNGV-LTGIVLSYHPLD-----EGGKQLSPFLRD 972
Qy 575 PVL 577
Db 973 PEL 975

```

RESULT 84

ID	KAC_HUMAN	STANDARD;	PRT;	106 AA.
AC	P01834;			
DT	21-JUL-1986 (Rel. 01, Created)			
DT	21-JUL-1986 (Rel. 01, Last sequence update)			
DT	15-MAR-2004 (Rel. 43, Last annotation update)			
DE	Ig kappa chain C region.			
GN	IGKC.			
OS	Homo sapiens (Human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.			
OX	NCBI_Taxid=9606;			
RN	[1]			
RP	SEQUENCE (MYELOMA PROTEIN EU).			
RX	MEDLINE=71064023; PubMed=5489770;			
RA	Gotlieb P.D., Cunningham B.A., Rutishauser U., Edelman G.M.;			
RT	"The covalent structure of a human gamma G-immunoglobulin. VI. Amino			
RT	acid sequence of the light chain."			
RL	Biochemistry 9:3155-3161(1970).			
RN	[2]			
RP	DISULFIDE BONDS.			
RX	MEDLINE=71064027; PubMed=4923144;			
RA	Gall W.E., Edelman G.M.;			
RT	"The covalent structure of a human gamma G-immunoglobulin. X.			
RT	intrachain disulfide bonds."			
RL	Biochemistry 9:3188-3196(1970).			
RN	[3]			
RP	SEQUENCE (BENCE-JONES PROTEIN TI).			
RX	MEDLINE=72188439; PubMed=5027703;			



RA Suter L., Barnikol H.U., Watanabe S., Hilschmann N.;  
 RT "Rule of antibody structure. The primary structure of a monoclonal  
 RT immunoglobulin L-chain of kappa-type, subgroup 3 (Bence-Jones protein  
 RT T1). IV. The complete amino acid sequence and its significance for  
 RT the mechanism of antibody production.";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 353:189-208 (1972).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=81042304; PubMed=6775818;  
 RA Hietzer P.A., Max E.E., Seidman J.G., Maizel J.V. Jr., Leder P.;  
 RT "Cloned human and mouse kappa immunoglobulin constant and J region  
 RT genes conserve homology in functional segments.";  
 RL Cell 22:197-207 (1980).  
 RN [5]  
 RP SEQUENCE (BENCE-JONES PROTEIN ROY).  
 RA Hilschmann N., Barnikol H.U., Hees M., Langer B., Ponstingl H.,  
 RL Steinmetz-Kayne M., Suter L., Watanabe S.;  
 RL (in) Frank F., Shugar D. (eds.);  
 RL Gamma globulins: structure and function, pp.57-74, Academic Press,  
 RL New York (1969).  
 RN [6]  
 RP SEQUENCE (BENCE-JONES PROTEIN CUM).  
 RX MEDLINE=68242259; PubMed=5586923;  
 RA Hilschmann N.;  
 RT "The complete amino acid sequence of Bence Jones protein Cum (kappa-  
 RT type).";  
 RL Hoppe-Seyler's Z. Physiol. Chem. 348:1718-1722 (1967).  
 RN [7]  
 RP SEQUENCE (BENCE-JONES PROTEIN AG).  
 RX MEDLINE=69234734; PubMed=4893682;  
 RA Tiliari K., Shindou T., Putnam F.W.;  
 RT "The amino acid sequence of a kappa type Bence-Jones protein. 3. The  
 RT complete sequence and the location of the disulfide bridges.";  
 RL J. Biol. Chem. 244:3550-3560 (1969).  
 RN [8]  
 RP SEQUENCE (WALDENSTROM'S MACROGLOBULIN OU).  
 RX MEDLINE=70201507; PubMed=5447531;  
 RA Kohler H., Shlimizu A., Paul C., Putnam F.W.;  
 RT "Macroglobulin structure: variable sequence of light and heavy  
 RT chains.";  
 RL Science 169:56-59 (1970).  
 RN [9]  
 RP SEQUENCE OF 1-33; 38-41 AND 62-80.  
 RC TISSUE=Abdominal adipose tissue;  
 RX MEDLINE=98249779; PubMed=9588180;  
 RA Olsen K.E., Sletten K., Westermarck P.;  
 RT "Extended analysis of AL-amyloid protein from abdominal wall  
 RT subcutaneous fat biopsy: kappa IV immunoglobulin light chain.";  
 RL Biochem. Biophys. Res. Commun. 245:713-716 (1998).  
 CC [1- MISCELLANEOUS: The EU sequence has the INV (3) alloypic marker,  
 CC 45-Ala and 83-Val. The ROY sequence has the INV (1,2) alloypic  
 CC marker, 45-Ala and 83-Leu.  
 CC [1- SIMILARITY: Contains 1 immunoglobulin-like domain.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 CC EMBL: J00241; AAA58989.1; -;  
 DR EMBL: V00557; CAA23823.1; -;  
 DR PIR: B90562; K3HU.  
 DR PDB: 1D5B; 03-FEB-00.  
 DR PDB: 1D5I; 09-FEB-00.  
 DR PDB: 1D6V; 04-OCT-00.  
 DR PDB: 1HEZ; 18-JUL-03.  
 DR PDB: 1HKU; 12-MAR-97.  
 DR PDB: 1I7Z; 08-AUG-01.  
 DR PDB: 1MIM; 15-MAY-97.  
 DR Genew; HGNC:5716; IGKC.

DR MIM; 147200; -;  
 DR GO: GO:0003823; F:antigen binding; NAS.  
 DR GO: GO:0006955; P:immune response; NAS.  
 DR InterPro: IPR007110; Ig-like.  
 DR InterPro: IPR003597; Ig-cl.  
 DR InterPro: IPR003006; Ig\_MHC.  
 DR Pfam: PF00047; Ig; 1.  
 DR SMART: SM00407; Igcl; 1.  
 DR PROSITE: PS00835; IG\_LIKE; 1.  
 DR PROSITE: PS00290; IG\_MHC; 1.  
 KW Immunoglobulin domain; Immunoglobulin C region; 3D-structure.  
 FT NON TER 1 102 IG-LIKE.  
 FT DOMAIN 5 102  
 FT DISULFID 26 86  
 FT DISULFID 106 106  
 FT VARIANT 83 83  
 FT CONFLICT 14 14 INTERCHAIN (WITH A HEAVY CHAIN).  
 FT CONFLICT 57 57 D->N (IN REF. 7 AND 8).  
 FT CONFLICT 57 57 E->Q (IN REF. 5 AND 6).  
 SQ SEQUENCE 106 AA; 11609 MW; 51984D1FDD372CE8 CRC64;  
 Query Match 4.3%; Score 148; DB 1; Length 106;  
 Best Local Similarity 29.5%; Pred. No. 0.004;  
 Matches 28; Conservative 24; Mismatches 41; Indels 2; Gaps 1;  
 Oy 525 PoyvTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEMSNGL--QENNYKKTPTPLDSDGS 582  
 Db 5 PSVTFPPSPDQLGSGTASVCLINNFYPRAKQWKVDNALQSGNSQESVTEQDSKDST 64  
 Oy 583 FFYLSKLTVDKSRMQGNGVFCSVYHBAHNNHYTQ 617  
 Db 65 YSLSTLTLLKADYKKYVACVETHGGLSSPTVK 99  
 RESULT 85  
 VCALL MOUSE STANDARD; PRT; 729 AA.  
 AC P29533;  
 DT 01-APR-1993 (Rel. 25, Created)  
 DT 01-APR-1993 (Rel. 25, Last sequence update)  
 DT 10-OCT-2003 (Rel. 42, Last annotation update)  
 DE Vascular cell adhesion protein 1 precursor (V-CAM 1).  
 GN VCAM1 OR VCAM-1.  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=FVB; TISSUE=Lung;  
 RX MEDLINE=92181437; PubMed=1371918;  
 RA Heesion C., Moy P., Tizard R., Chisholm P., Williams C., Wysk M.,  
 RA Burkly L., Miyake K., Kincade P., Lobb R.;  
 RT "Cloning of murine and rat vascular cell adhesion molecule-1.";  
 RL Biochem. Biophys. Res. Commun. 183:163-169 (1992).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Lymph node;  
 RX MEDLINE=93246254; PubMed=7683304;  
 RA Arai M., Arai K., Vassalli P.;  
 RT "Cloning and sequencing of mouse VCAM-1 cDNA.";  
 RL Gene 126:261-264 (1993).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC STRAIN=129; TISSUE=Embryo;  
 RX MEDLINE=94117008; PubMed=7507076;  
 RA Cybulsky M.I., Allan-Motamed M., Collins T.;  
 RT "Structure of the murine VCAM1 gene.";  
 RL Genomics 18:387-391 (1993).  
 RN [4]  
 RP SEQUENCE OF 1-693 FROM N.A. (ISOFORM 1).  
 RC STRAIN=NIH Swiss, and 129/Sv;  
 RA Kumar A.G., Dai Y.X., Kozak C.A., Mims M.P., Gotto A.M. Jr.,

RA Ballantyne C.M.;  
 RL Submitted (AUG-1994) to the EMBL/Genbank/DBJ databases.  
 RN [5]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=FVB; TISSUE=Lung;  
 RX MEDLINE=93232042; PubMed=7682556;  
 RA Moy P., Lobb R., Tizard R., Olson D., Hession C.;  
 RT "Cloning of an inflammation-specific phosphatidyl inositol-linked  
 RL form of murine vascular cell adhesion molecule-1.";  
 RN J. Biol. Chem. 268:8835-8841(1993).  
 RN [6]  
 RP SEQUENCE FROM N.A. (ISOFORM 2).  
 RC STRAIN=C57BL/6; TISSUE=Liver;  
 RX MEDLINE=95015899; PubMed=7523515;  
 RA Kumar A.G., Dai X.Y., Kozak C.A., Mims M.P., Gotto A.M.,  
 RA Ballantyne C.M.;  
 RT "Murine VCAM-1. Molecular cloning, mapping, and analysis of a  
 RL truncated form.";  
 RN J. Immunol. 153:4088-4098(1994).  
 RN [7]  
 RP SEQUENCE OF 311-345 FROM N.A. (ISOFORM 2).  
 RC STRAIN=FVB/N; TISSUE=Kidney;  
 RX MEDLINE=93317595; PubMed=7687058;  
 RA Terry R.W., Kwee L., Levine J.F., Labow M.A.;  
 RT "Cytokine induction of an alternatively spliced murine vascular cell  
 RL adhesion molecule (VCAM) mRNA encoding a  
 RT glycosylphosphatidylinositol-anchored VCAM protein.";  
 RN Proc. Natl. Acad. Sci. U.S.A. 90:5919-5923(1993).  
 RN [8]  
 RP SEQUENCE OF 1-21 FROM N.A.  
 RC TISSUE=Endothelial cells;  
 RA Korenaga R., Ando J., Tsuboi H., Kamiya A.;  
 RL Submitted (DEC-1995) to the EMBL/Genbank/DBJ databases.  
 CC -1- FUNCTION: IMPORTANT IN CELL-CELL RECOGNITION. APPEARS TO FUNCTION  
 CC IN LEUKOCYTE-ENDOTHELIAL CELL ADHESION. INTERACTS WITH THE BETA-1  
 CC INTEGRIN VIA4 ON LEUKOCYTES, AND MEDIATES BOTH ADHESION AND SIGNAL  
 CC TRANSDUCTION. THE VCAM1/VLA4 INTERACTION MAY PLAY A  
 CC PATHOPHYSIOLOGIC ROLE BOTH IN IMMUNE RESPONSES AND IN LEUKOCYTE  
 CC EMIGRATION TO SITES OF INFLAMMATION.  
 CC -1- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN (LONG ISOFORM) OR  
 CC ATTACHED TO THE MEMBRANE BY A GPI-ANCHOR (SHORT ISOFORM).  
 CC -1- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=1; Synonym=Long;  
 CC IsoId=P29533-1; Sequence=Displayed;  
 CC Name=2; Synonym=Short;  
 CC IsoId=P29533-2; Sequence=VSP 002581, VSP 002582;  
 CC -1- TISSUE SPECIFICITY: Expressed on inflamed vascular endothelium, as  
 CC well as on macrophage-like and dendritic cell types in both normal  
 CC and inflamed tissue.  
 CC -1- PTM: The GPI-anchor is located on position 319 of isoform 2.  
 CC -1- SIMILARITY: Contains 7 immunoglobulin-like C2-type domains.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see <http://www.isb-slb.ch/announce/>  
 CC or send an email to [license@isb-slb.ch](mailto:license@isb-slb.ch)).  
 CC -----  
 DR EMBL; M84487; AAA40545.1; -;  
 DR EMBL; X67783; CAA47989.1; -;  
 DR EMBL; L22355; AAA16921.1; -;  
 DR EMBL; L22301; AAA16921.1; JOINED.  
 DR EMBL; L22349; AAA16921.1; JOINED.  
 DR EMBL; L22350; AAA16921.1; JOINED.  
 DR EMBL; L22351; AAA16921.1; JOINED.  
 DR EMBL; L22352; AAA16921.1; JOINED.  
 DR EMBL; L22353; AAA16921.1; JOINED.  
 DR EMBL; L22354; AAA16921.1; JOINED.  
 DR EMBL; L22350; AAA16920.1; -;  
 DR EMBL; L22301; AAA16920.1; JOINED.

[illegible]

```

Db 271 SGNATLTLIIMARMEDSGVYVEGVNLIGRDAEVELVQEKPFIVDISPGSQAUAQVGS 330
Oy 139 LTLTLESPGSSPSVQCRSPGKNIQ-----GKTLVSQLELDGSGTWTCTYLNQX 191
Db 331 VLTICAAICGDSPSRMTQDSDPLANGVYRNEGAKSTIVLVSQGFEDHSLCAVTCIOR 390
Oy 192 KVEFKIDIVLAFQKASSIVYKKEGQVFSFPLA-----FTVEKL 232
Db 391 TLEKRTQVEVYSF-----PEDPVIKMSGPLVHGRPVTVNCTPVNVYFPDHLIELL 441
Oy 233 TGSGELMWQAEARASSKSMI-TFDLKNKESVYKRYTOPKQ-NGKYL-----PLHLTIPO 286
Db 442 KG-----ETTLMKKYLLEEMGKSLKLETLTTPFPTIEDGKSLVCLARLHSGEME 493
Oy 287 ALPOYASGSGNLTALFAKTKLHGVNLVWNRATOLQK-----NLTCVWGPTSPKLMSTL 342
Db 494 SEPKROSVQPLVYVAVP-----KTTTWSPSPLELGGSPVNLCCSDGIPAPKILMSR 548
Oy 343 KLENKAKVSKREKFWVJLN--PEAGNMOCILSDSG--QVLESNIKVLPTWSTPVP 396
Db 549 QLNNGELPLSENTLTFTMSKTRDSDGIYVCEGINAIGSKSVELLIQV----- 598
Oy 397 PAPEKSCDKTHTCELLIGSPSVLFP--KPKDTLMISRT---PEVTCV 442
Db 599 --SPKDIQLT-----VFPKSVYEGDVIISCTCGNVFETWILL 635

RESULT 86
ID LAC1_RAT STANDARD; PRT; 104 AA.
AC P20766;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DE 10-OCT-2003 (Rel. 42, Last annotation update)
DE 19_Lambda-1 chain C region.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87305594; PubMed=3114047;
RA Steen M.L., Hellman L., Petersson U.;
RT "The immunoglobulin lambda locus in rat consists of two C lambda
genes and a single V lambda gene.";
RL R.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M22820; AAA41419.1; ALT_INIT.
DR HSSP; P01842; 7EAB.
DR InterPro; IPR007110; IG-1-like.
DR InterPro; IPR003597; IG_C1.
DR InterPro; IPR003006; IG_MHC.
DR Pfam; PF00047; Ig_1.
DR SMART; SM00407; IgC1.1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON TER 1
FT DOMAIN 1
FT DISULFID 6 99 IG-LIKE.
FT DISULFID 27 85
FT DISULFID 103 103 INTERCHAIN (WITH HEAVY CHAIN).
SQ SEQUENCE 104 AA; 11565 MW; C8B71811F4B8978A CRC64;
Query Match 4.3%; Score 146; DB 1; Length 104;

```

```

Best Local Similarity 33.0%; Pred. No. 0.0052;
Matches 34; Conservative 19; Mismatches 44; Indels 6; Gaps 4;
Oy 521 QPR-EPQVYTLPEPRDLTKNQVSLTCLVNGFSPSDIAVMESNGQ-ENNYKTPPVLD 578
Db 1 QPKATPSVTLFPPSESLDKATLVCMVIDFPGVATVWKADGITGQVETTP- 59
Oy 579 SDGSFPLYSKLTVDKSRWQGNVSCVMEALHNHYTKSL 621
Db 60 QNNKYATSYLLTKAKMETHSNVSCVYTH--ENVYKSL 99

RESULT 87
SRB2_HUMAN STANDARD; PRT; 387 AA.
ID SRB2_HUMAN
AC Q9P1R8; Q9WMA5; Q9NQK8;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Signal-regulatory protein beta-2 precursor (SIRP-beta-2) (SIRP-b2).
GN SIRPB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORM 2).
RC TISSUE=Placenta;
RX MEDLINE=21036165; PubMed=11185750;
RA Ichigotani Y., Matuda S., Machida K., Oshima K., Iwamoto T.,
RA Yamaki K., Hayakawa T., Haseguchi M.;
RT "Molecular cloning of a novel human gene (SIRP-B2) which encodes a new
member of the SIRP/SHB-1 protein family.";
RL J. Hum. Genet. 45:378-382(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=21638749; PubMed=11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhand P.D., Dunn M.,
RA Ellington A.G., Griffiths J.A., Frazer A., French L., Garner P.,
RA Grafham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Leharvalho M.H., Levesha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McDonach L.J., McKay K., McMurray A.A.,
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillips B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,
RA Rice C.M., Ross M.T., Scott C.B., Senta H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sutcliffe J.E.,
RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,
RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitlaker P., Willey D.L., Williams L., Williams S.A.,
RA Wilmshurst L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT "The DNA sequence and comparative analysis of human chromosome 20.";
RL Nature 414:865-871(2001).
RN [3]
RP SEQUENCE FROM N.A. (ISOFORM 3).
RC TISSUE=Lung;
RX MEDLINE=22386257; PubMed=12477932;
RA Strauberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shenner C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heich F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,

```

FT	CARBOHYD	309	309	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	CARBOHYD	317	317	N-LINKED (GLCNAC. . .)	(POTENTIAL).
FT	VARSPLIC	1	33	Missing (in isoform 2).	
FT	VARSPLIC	144	360	/FtId=VSP_007027.	
FT	VARSPLIC	263	263	Missing (in isoform 3).	
FT	CONFLICT	286	286	/FtId=VSP_007028.	
FT	CONFLICT	387 AA,	42495 MM;	A -> V (IN REF. 2).	
FT	SEQUENCE	387 AA,	42495 MM;	L -> S (IN REF. 2).	
FT	SEQUENCE	387 AA,	42495 MM;	F7F20C9F8E6DE64B CRC64;	
Query Match		4.3%;	Score 146;	DB 1;	Length 387;
Best Local Similarity		24.8%;	Pred. No. 0.029;		
Matches	57,	Conservative 37;	Mismatches 114;	Indels 22;	Gaps 8;
OY	412 ELUGSEPV-FLFPKPKDTLMIS----	RTEYVCVVVDVSH--	EDPEVKENWYDVGEVH	464	
Db	131 EFKSGPTENALAKAPSAVPLGPAARTTPEHTVSFTCESHGSPRDIITLKFENGNELS	190			
OY	465 NAKTKPREQGYNSYRVSVSLTVLHDQMLNGKEYCKCVSKALPA-PIEKT--ISKAKGQ	521			
Db	191 DFGNVNDPTGOSAVYSIRSTARVVLDPMDVRSCVCEVAHVTLGGDPLRGNTANLSAIRV	250			
OY	522 PREFOVLTLPSSDELTKNOVSTLCVKGFPSPSDIAVEMESNQPENNYKTTTPVLDSDG	581			
Db	251 PRTLEV-TQQPMR---AGNQNVTTCQYRKRYPSQLGTWTJLENGVNCQRETASTLTEKDQ	306			
OY	582 SFPLYSLCTVYDKSRMGQGVFCSGVMHEALHNHYTKSLSLSPGLQDET	631			
Db	307 TWMTSWFLVNISPDRDDVLTCCQVKADG-----QLAVSKRLALEVT	348			
RESULT 88					
ID	LAC_CHICK	STANDARD;	PRT;	103 AA.	
AC	P20763;				
DT	01-FEB-1991 (Rel. 17, Created)				
DT	01-FEB-1991 (Rel. 17, Last sequence update)				
DT	10-OCT-2003 (Rel. 42, Last annotation update)				
DS	19_Lambda chain C region.				
OS	Gallus gallus (chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
CC	Gallus.				
CC	NCBI_TaxID=9031;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RX	MEDLINE=87128480; PubMed=1107981;				
RA	Parvati R., Ziv E., Lientner F., Tel-Or S., Bursstein Y., Schechter I.;				
RT	"Analyses of chicken immunoglobulin light chain cDNA clones indicate				
RL	a few germline y lambda genes and allotypes of the C lambda locus.";				
EMO	J. 6:197-102(1987).				
CC	-1 SIMILARITY: Contains 1 immunoglobulin-like domain.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.ebi.ac.uk/announcement/				
CC	or send an email to license@ebi.ac.uk).				
CC	-----				
DR	EMBL; X04768; CAZ28461.1; -.				
DR	PIR; B26167; B26167.				
DR	HSSP; P01842; P01842.				
DR	InterPro; IPRO07110; IG-like.				
DR	InterPro; IPR003597; IG cl.				
DR	InterPro; IPR003006; IG_MHC.				
DR	Pfam; PF00047; Ig; 1.				
DR	SMART; SM00407; IgC1; 1.				
DR	PROSITE; PS50835; IG_LIKE; 1.				
DR	PROSITE; PS00290; IG_MHC; 1.				
KM	Immunoglobulin domain; Immunoglobulin C region.				
FT	NON_TER	1			

```
FT DOMAIN 6 99 IG-LIKE.
FT DISULFID 28 85 INTERCHAIN (WITH HEAVY CHAIN).
FT DISULFID 103 103 N -> D.
FT VARIANT 90 90
SO SEQUENCE 103 AA; 11361 MW; 77BF341B511B91B2 CRC64;

Query Match 4.3%; Score 145.5; DB 1; Length 103;
Best Local Similarity 32.2%; Pred. No. 0.0056;
Matches 29; Conservative 19; Mismatches 39; Indels 3; Gaps 3;

Oy 521 QPR-EPQVYTLPPSRDELTK-NQVSLTCLVKGFPSDIAVEMESNGOPENNYKTTTPYLD 578
Db 1 QPKVAPRTLPFPSPKEELINAEATKATLVCLINDFYPSPVTVMDVIGSTRSG-ETTAPOQ 59
Oy 579 SDGSFPLYSKLTVDKSRMOQGNVFCSSVME 608
Db 60 SNSQYMASSYLSLSDMSWSHETVCRAVTH 89

RESULT 89
KACB_RAT
ID_KACB_RAT STANDARD; PRT; 106 AA.
AC P01835;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Ig kappa chain C region, B allele.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Jouvalin;
RX MEDLINE=82082587; PubMed=62733908;
RA Sheppard H.W., Gutman G.A.;
RT "Allelic forms of rat kappa chain genes: evidence for strong
RL selection at the level of nucleotide sequence.";
Proc. Natl. Acad. Sci. U.S.A. 78:7064-7068(1981).
RN [2]
RP SEQUENCE (BENCE-JONES PROTEIN S211).
RX MEDLINE=75212238; PubMed=807630;
RA Starace V., Querintjean P.;
RT "The primary structure of a rat kappa Bence Jones protein:
RL phylogenetic relationships of V- and C-region genes.";
J. Immunol. 115:59-62(1975).
CC -1 SIMILARITY: Contains 1 immunoglobulin-like domain.
DR HSP; A93901; K1RTB.
DR HSP; P01842; 2MCG.
DR InterPro; IPR000710; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR Pfam; PF00047; Ig1.
DR SMART; SM00407; IGc1.1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region.
FT NON_TER 1 1
FT DOMAIN 5 102 IG-LIKE.
FT DISULFID 26 86 INTERCHAIN (WITH A HEAVY CHAIN).
FT DISULFID 106 106
FT CONFLICT 2 2 D -> N (IN REF. 2).
FT CONFLICT 30 30 N -> K (IN REF. 2).
FT CONFLICT 48 48 MISSING (IN REF. 2).
FT CONFLICT 79 79 E -> Q (IN REF. 2).
FT CONFLICT 87 87 E -> Q (IN REF. 2).
FT CONFLICT 98 98 V -> Vw (IN REF. 2).
FT CONFLICT 100 100 S -> N (IN REF. 2).
SO SEQUENCE 106 AA; 11601 MW; 4CFA7CA820D1CA36 CRC64;

Query Match 4.2%; Score 145; DB 1; Length 106;
Best Local Similarity 31.5%; Pred. No. 0.0062;
Matches 29; Conservative 22; Mismatches 29; Indels 12; Gaps 2;
```

```
Oy 525 PQVYTLPPSRDELTKNQVSLTCLVKGFPSDIAVEMESNGOPENNYKTTTPYLD 578
Db 5 PTYSIFPSPSTEQLTGASVYVCLMNNFPRDISVKMKIDGTERD-----GVLDSTVD 59
Oy 580 --DGSPFLYSKLTVDKSRMOQGNVFCSSVME 609
Db 60 SKDSTYSMSSTLSLTAKDYESHNLVTCVEVHK 91

RESULT 90
HB2C_PIG
ID_HB2C_PIG STANDARD; PRT; 261 AA.
AC P15982;
DT 01-APR-1990 (Rel. 14, Created)
DT 01-APR-1990 (Rel. 14, Last sequence update)
DT 15-JUN-1999 (Rel. 38, Last annotation update)
DE SLA class II histocompatibility antigen, DQ haplotype C beta chain
DE precursor.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_Taxid=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=90361905; PubMed=2391424;
RA Gustafsson K., Leguenn C., Hirsch F., Germana S., Pratt K.,
RA Sachs D.H.;
RT "Class II genes of miniature swine, IV. Characterization and
RL expression of two allelic class II DQB cDNA clones.";
J. Immunol. 145:1946-1951(1990).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcements
CC or send an email to license@ebi.ac.uk).
CC -----
DR EMBL; M31497; AAA31084.1; -
DR EMBL; M32117; AAA53110.1; -
DR PIR; A60404; A60404.
DR HSP; P13760; 2SEB.
DR InterPro; IPR007110; Ig-like.
DR InterPro; IPR003597; Ig_c1.
DR InterPro; IPR003006; Ig_MHC.
DR InterPro; IPR000353; MHC_II_beta.
DR Pfam; PF00047; Ig1.
DR Pfam; PF00969; MHC_II_beta; 1.
DR ProDom; PD000328; MHC_II_beta; 1.
DR SMART; SM00407; IGc1.1.
DR PROSITE; PS50835; IG_LIKE; 1.
DR PROSITE; PS00290; IG_MHC; 1.
KW MHC II; Transmembrane; Glycoprotein; Signal.
FT SIGNAL 1 31
FT CHAIN 32 261
FT FT 32 261 SLA CLASS II HISTOCOMPATIBILITY ANTIGEN,
FT FT 32 261 DO HAPLOTYPE C BETA CHAIN.
FT DOMAIN 127 126 EXTRACELLULAR BETA-1.
FT DOMAIN 221 230 EXTRACELLULAR BETA-2.
FT TRANSMEM 231 251 CONNECTING PEPTIDE.
FT DOMAIN 252 251 CYTOPLASMIC TAIL.
FT DISULFID 47 111 BY SIMILARITY.
FT DISULFID 149 205 BY SIMILARITY.
FT CARBOHYD 51 51 N-LINKED (GLCNAC...) (POTENTIAL).
SO SEQUENCE 261 AA; 29550 MW; AA9581F2A3B1969D CRC64;

Query Match 4.2%; Score 145; DB 1; Length 261;
Best Local Similarity 24.7%; Pred. No. 0.02;
Matches 49; Conservative 36; Mismatches 93; Indels 20; Gaps 6;
```

```

DB      36 PDGFVQFK-GECEYFNGCVRGAVARIYNGEHLRFSDVGEFAVTPPLGRPEADSN 94
QY      477 STYRVSVLTVLHQMNLNGEKYCKVSNKALPAPIEKITSKAGQREPOVYTLPPSRDE 536
DB      95 SQDVLQKRA-EVDVRCCKNYQIE-----EGTTLQRAVQF---TWIISPSKAE 139
QY      537 LTNGQVSLTCLVGVGFPSDIAVWESNGQPEN-NYKTTTPVLDSDSFLYSLTLVDKSR 595
DB      140 ALMHNHLVCAVVDFFPSQVKVQMFNNGQETAGVSTPLIRNGD---WTYQVLVLENN 196
QY      596 MOQNVFSCSVMEALHN 613
DB      197 LQRGDVYTCRVEHSION 214

RESULT 91
SHS1_RAT STANDARD: PRT; 509 AA.
ID SHS1_RAT 008951; 070426; 09QW15;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DE Protein-tyrosine phosphatase non-receptor type substructure 1 precursor
DE (SHP substrate-1) (SHPS-1) (inhibitory receptor SHPS-1) (Signal-
DE regulatory protein alpha-1) (SIRP-alpha-1) (Brain Ig-like molecule
DE with tyrosine-based activation motif) (Bit) (Macrophage fusion
DE receptor) (Macrophage membrane protein MFP150).
GN PRPNS1 OR SHPS1 OR SIRP OR BIT OR MFR.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
[1]
SEQUENCE FROM N.A., SEQUENCE OF 42-60; 68-91; 128-137; 150-158;
RP 174-189; 192-202; 204-212; 218-237; 259-270; 279-282; 405-415 AND
RP 446-453. N-GLYCOSYLATION, PHOSPHORYLATION ON TYROSINE RESIDUES, AND
RP INTERACTIONS WITH PTPN6 AND PTPN11.
RC TISSUE=Fetal Fibroblast; PubMed=8943344;
RX MEDLINE=97098667; PubMed=8943344;
RA Fujioaka Y., Matozaki T., Noguchi T., Iwamatsu A., Yamao T.,
RA Takahashi N., Tsuda M., Takada T., Kasuga M.;
RT "A novel membrane glycoprotein, SHPS-1, that binds the SH2-domain-
RT containing protein tyrosine phosphatase SHP-2 in response to mitogens
RT and cell adhesion.";
RL Mol. Cell. Biol. 16:6887-6899(1996).
[2]
SEQUENCE FROM N.A., SEQUENCE OF 32-48 AND 446-453, FUNCTION, AND
RP PHOSPHORYLATION ON TYROSINE RESIDUES.
RC STRAIN=Sprague-Dawley; TISSUE=Brain;
RX MEDLINE=97415431; PubMed=9271230;
RA Sano S.-I., Ohnishi H., Omori A., Hasegawa J., Kubota M.;
RT "BIT, an immune antigen receptor-like molecule in the brain.";
RL FEBS Lett. 411:327-334(1997).
[3]
SEQUENCE FROM N.A., SEQUENCE OF 99-107; 128-149; 192-217; 405-417;
RP 419-429; 446-467 AND 496-506. N-GLYCOSYLATION, AND TISSUE SPECIFICITY.
RC STRAIN=Fischer 344; TISSUE=Macrophage;
RX MEDLINE=98449911; PubMed=9774638;
RA Segnario C., Sterling H., Beckers C., Kobayashi R., Solimena M.,
RA Ullu E., Vignery A.;
RT "MFR, a putative receptor mediating the fusion of macrophages.";
RL Mol. Cell. Biol. 18:6213-6223(1998).
[4]
SEQUENCE OF 1-419 FROM N.A., AND TISSUE SPECIFICITY.
RC STRAIN=MG/Rij; TISSUE=Alveolar macrophage;
RX MEDLINE=98375871; PubMed=9712053;
RA Adams S., van der Laan L.J.W., Vernon-Wilson E.,
RA Renardel de Lavalette C., Doepf E.A., Dijkstra C.D., Simmons D.L.,
RT "Signal-regulatory protein is selectively expressed by myeloid and
RT neuronal cells.";
RL J. Immunol. 161:1853-1859(1998).

```

```

RN      [5]
RP PHOSPHORYLATION IN RESPONSE TO EGF, AND INTERACTION WITH PTPN11.
RX MEDLINE=98008865; PubMed=9344856; Fujioaka Y., Yamao T., Takada T.,
RX Ochi F., Matozaki T., Noguchi T.,
RA Tsuda M., Takada H., Fukunaga K., Okabayashi Y., Kasuga M.;
RT "Epidermal growth factor stimulates the tyrosine phosphorylation of
RT SHPS-1 and association of SHPS-1 with SHP-2, a SH2 domain-containing
RT protein tyrosine phosphatase.";
RL Biochem. Biophys. Res. Commun. 239:483-487(1997).
RN      [6]
RP PHOSPHORYLATION OF TYR-477 AND TYR-501, AND MUTAGENESIS OF TYR-436;
RX TYR-460; TYR-477 AND TYR-501.
RX MEDLINE=98204923; PubMed=9535915;
RA Takada T., Matozaki T., Takada H., Fukunaga K., Noguchi T.,
RA Fujioaka Y., Okazaki I., Tsuda M., Yamao T., Ochi F., Kasuga M.;
RT "Roles of the complex formation of SHPS-1 with SHP-2 in
RT insulin-stimulated mitogen-activated protein kinase activation.";
RL J. Biol. Chem. 273:9234-9242(1998).
CC -1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts
CC as docking protein and induces translocation of PTPN6,
CC PTPN11 and other binding partners from the cytosol to the
CC plasma membrane. Supports adhesion of cerebellar neurons, neurite
CC outgrowth and glial cell attachment. May play a key role in
CC intracellular signaling during synaptogenesis and in synaptic
CC function. Involved in the negative regulation of receptor tyrosine
CC kinase-coupled cellular responses induced by cell adhesion, growth
CC factors or insulin. Mediates negative regulation of phagocytosis,
CC mast cell activation and dendritic cell activation. CD47 binding
CC prevents maturation of immature dendritic cells and inhibits
CC cytokine production by mature dendritic cells. May play a role in
CC the release of nitric oxide by macrophages (By similarity).
CC -1- SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in
CC macrophages, where it primarily binds PTPN6. Binds GRB2 in vitro.
CC Binds FGR. Binds JAK2 irrespective of its phosphorylation status
CC and forms a stable complex. Binds SCAPI and/or SCAPI2. The
CC resulting complex recruits Fyb. Binds PRK2B (By similarity).
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.
CC -1- TISSUE SPECIFICITY: Highly expressed in brain, spleen, lung, liver
CC and kidney. Detected at lower levels in heart. Highly expressed in
CC alveolar and peritoneal macrophages, and at lower levels in
CC dendritic cells.
CC -1- PTM: N-glycosylated.
CC -1- PTM: Phosphorylated on tyrosine residues in response to insulin,
CC cell adhesion or epidermal growth factors. Dephosphorylated by
CC PTPN11.
CC -1- SIMILARITY: Contains 2 immunoglobulin-like C1-type domains.
CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL, DB5133; BA012734.1; -
DR EMBL, D38468; AAC20368.1; -
DR EMBL, U62328; AAC68478.1; -
DR EMBL, AF055065; AAC18089.1; -
DR HSSP, P01703; 7EAB.
DR InterPro: IPR007110; Ig-like.
DR InterPro: IPR003587; Ig_C1.
DR InterPro: IPR003006; Ig_MHC.
DR Pfam, PF00047; Ig1; 3.
DR SMART, SM00407; IG1; 2.
DR PROSITE, PS00290; IG_MHC; 3.
DR Repeat, Signal; Transmembrane, Immunoglobulin domain; SH3-binding;
KW Glycoprotein; Phosphorylation.
FT SIGNAL 1 31
FT CHAIN 32 509
FT TYPE SUBSTRATE 1.

```

```

FT DOMAIN 28 373 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 374 394 POTENTIAL.
FT DOMAIN 395 509 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 32 138 IG-LIKE V-TYPE.
FT DOMAIN 150 248 IG-LIKE C1-TYPE 1.
FT DOMAIN 235 349 IG-LIKE C1-TYPE 2.
FT DISULFID 55 122 POTENTIAL.
FT DISULFID 172 229 POTENTIAL.
FT DISULFID 274 332 POTENTIAL.
FT SITE 436 439 SH2-BINDING (POTENTIAL).
FT SITE 446 451 SH3-BINDING (POTENTIAL).
FT SITE 460 463 SH2-BINDING (POTENTIAL).
FT SITE 477 480 SH2-BINDING (POTENTIAL).
FT SITE 501 504 SH2-BINDING (POTENTIAL).
FT MOD_RES 436 436 PHOSPHORYLATION (BY TYR-KINASES) (POTENTIAL).
FT MOD_RES 460 460 PHOSPHORYLATION (BY TYR-KINASES) (POTENTIAL).
FT MOD_RES 477 477 PHOSPHORYLATION (BY TYR-KINASES) (POTENTIAL).
FT MOD_RES 501 501 PHOSPHORYLATION (BY TYR-KINASES) (POTENTIAL).
FT CARBOHYD 54 54 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 93 93 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 169 169 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 181 181 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 205 205 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 209 209 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 242 242 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 246 246 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 271 271 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 293 293 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 312 312 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 345 345 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 436 436 Y->F: ABOLISHES TYROSINE PHOSPHORYLATION AND PTPL1 BINDING; WHEN ASSOCIATED WITH F-436, F-477 AND F-501.
FT MUTAGEN 460 460 Y->F: ABOLISHES TYROSINE PHOSPHORYLATION AND PTPL1 BINDING; WHEN ASSOCIATED WITH F-436, F-477 AND F-501.
FT MUTAGEN 477 477 Y->F: STRONGLY REDUCES INSULIN-INDUCED TYROSINE PHOSPHORYLATION AND PTPL1 BINDING. ABOLISHES TYROSINE PHOSPHORYLATION AND PTPL1 BINDING. WHEN ASSOCIATED WITH F-436, F-460 AND F-501.
FT MUTAGEN 501 501 Y->F: STRONGLY REDUCES INSULIN-INDUCED TYROSINE PHOSPHORYLATION AND PTPL1 BINDING. ABOLISHES TYROSINE PHOSPHORYLATION AND PTPL1 BINDING; WHEN ASSOCIATED WITH F-436, F-460 AND F-477.
FT CONFLICT 8 8 MISSING (IN REF. 4).
FT CONFLICT 10 10 F->I (IN REF. 3).
FT CONFLICT 25 25 S->C (IN REF. 4).
FT CONFLICT 58 58 KR->MP (IN REF. 3; AA SEQUENCE).
FT CONFLICT 99 100 G->A (IN REF. 2).
FT CONFLICT 162 162 D->N (IN REF. 3).
FT CONFLICT 189 189 N->L (IN REF. 3; AA SEQUENCE).
FT CONFLICT 205 205 N->G (IN REF. 3; AA SEQUENCE).
FT CONFLICT 209 209 G->F (IN REF. 3; AA SEQUENCE).
FT CONFLICT 405 405 E->P (IN REF. 3; AA SEQUENCE).
FT CONFLICT 416 421 NARE->EGON (IN REF. 3; AA SEQUENCE).
FT CONFLICT 418 421 R->E (IN REF. 3; AA SEQUENCE).
FT CONFLICT 450 450 MISSING (IN REF. 3; AA SEQUENCE).
FT CONFLICT 499 499 MISSING (IN REF. 3; AA SEQUENCE).
SO SEQUENCE 509 AA: 55690 MW: 581FE0A4D429F4 CRC64;
Query Match 4.2%; Score 145; DB 1; Length 509;
Best Local Similarity 21.5%; Pred. No. 0.048;
Matches 82; Conservativity 64; Mismatches 161; Indels 74; Gaps 16;

```

```

Qy 314 LVMRATOLQKNTLCEWGPSTSPKMLSLKLENKAKVSKREKRPVWNLNBEAGMOCILS 373
Db 92 TNSVDARK-RNNMD-----FSICISN-----VTPEDAGTYCVKF 125
Qy 374 DSGVTLIESNIK-----VLPWSTVPCPAPEPKSCDKHTCPBLGSGVFLPPPKKD 428
Db 126 QKGVSEDTREIKSGGGTTLVYLAKP---SSPEVSG-----PDSRSGP----- 164
Qy 429 TLMISREVTGVVDVSHEDPEYKFMWYDGVVHNAKTRPREQNSRYRVSVLTV- 487
Db 165 ----GQVNFCTKSYGSPNN--ITLKMKGKELSHLETTSKS--NVSINISSTVSVK 217
Qy 488 LHQDWLNKKEYKCVSKNALPA-PIEKTSKAKGQPREPOVYTLPSRDELTL-KNQVSLT 545
Db 218 LSPEDHSRIV-CEVAHVTLEGRPLNGTANFSNIRSP---TLKITQPLTPASQVNL 273
Qy 546 CLVGFYPSDIAVWESNGQPENNYKTPPVLDGSGFFLYSKLTYDKSRWQGNVFCSS 605
Db 274 CQVQKFPKALQNLWLENGMLSRDKREHPTDRDGTNYTSLFLVNSAHRBDVVFQC 333
Qy 606 VMHE---ALNNHTQKSLIS 623
Db 334 VERDSQPATENHTVRAFAHS 354
RESULT 92
NCM2_MOUSE STANDARD; PRT; 837 AA.
ID NCM2_MOUSE
AC 035136; 035962;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Neural cell adhesion molecule 2 precursor (N-CAM 2) (RB-8 neural cell adhesion molecule) (RAB12).
GN NCM2 OR OCAM OR RNCAM.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A. (ISOFORMS LONG AND SHORT).
RC STRAIN=BALB/c; TISSUE=Olfactory neuroepithelium;
RX MEDLINE=97368238; Pubmed=9221781;
RA Yoshitara Y., Kawasaki M., Tamada A., Fujita H., Hayashi H., Kagamiyama H., Mori K.;
RT "OCAM: A new member of the neural cell adhesion molecule family related to zone-to-zone projection of olfactory and vomeronasal axons."
RT J. Neurosci. 17:5830-5842(1997).
RL [2]
RN SEQUENCE FROM N.A. (ISOFORM SHORT).
RC STRAIN=C57BL/6J; TISSUE=Olfactory epithelium;
RX MEDLINE=97476194; Pubmed=9331170;
RA Alenius M., Bohm S.;
RT "Identification of a novel neural cell adhesion molecule-related gene with a potential role in selective axonal projection."
RT J. Biol. Chem. 272:26083-26086(1997).
CC -1- FUNCTION: May play important roles in selective fasciculation and zone-to-zone projection of the primary olfactory axons.
CC -1- SUBCELLULAR LOCATION: Type I membrane protein (long isoform) and attached to the membrane by a GPI-anchor (short isoform).
CC -1- ALTERNATIVE PRODUCTS:
Event=Alternative splicing; Named isoforms=2;
Name=Long;
IsoId=O35136-1; Sequence=Displayed;
CC -1- IsoId=O35136-2; Sequence=VSP_002590;
CC -1- TISSUE SPECIFICITY: Expressed in subsets of both olfactory and vomeronasal neurons in a zone-specific manner.
CC -1- SIMILARITY: Contains 5 immunoglobulin-like C2-type domains.
CC -1- SIMILARITY: Contains 2 fibronectin type III domains.
-----
This SWISS-Prot entry is copyright. It is produced through a collaboration

```

QY 3 RG-VPRFHLLVLDTALLPAPATGSKNVLLGKGPTEVLTETCTAGSOKSIOQHMKNSNOIKI 61  
Db 193 RGEIDPFDDITIVVVPAPIMMPQKSFNATARGSEMTLTCKASGSDPPTISMFRNG--KL 250  
QY 62 LGNGSGSLTGPSTLNDRADSRRLMDQGNFPLIKLKLIKEDSTYICEV----EYKE 116  
Db 251 IEENEKYLKG-----SNTELTVRNIIKNDGGSYVCKATKNAGBQKQ 293  
QY 117 E-VQLVFGLTANSDTHLLOGS-----LITLESPPGSSPVCC----- 155  
Db 294 AFLQVFEV-----QPHILQLKNETSENHGHTLVLCBAEGPVPBEIMTKBADGVMSSEG 346  
QY 156 -RSRPG-----KNIOGKTLVSQLELDDSGMTVTYIQ-----NOKKVEFKIDIVLAFQK 206

[illegible]



CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.ebi.ac.uk/announce/>  
CC or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).

CC EMBL: L25056; AAC38015.2; -  
DR PIR: I50478; I50478.  
DR HSSP: Q13740; 1KJC.  
DR InterPro: IPR007110; IG-like.  
DR InterPro: IPR003006; IG\_MHC.  
DR InterPro: IPR003596; IG\_V.  
DR Pfam: PF00047; Ig; 3.  
DR SMART: SM00406; IG; 1.  
DR PROSITE: PS00835; IG LIKE; 4.  
DR PROSITE: PS00290; IG MHC; 1.  
DR Signal: Developmental protein; Cell adhesion; Immunoglobulin domain;  
KW Repeat: Glycoprotein; Transmembrane.  
FT SIGNAL 1 22 POTENTIAL.  
FT CHAIN 23 555 CD166 ANTIGEN HOMOLOG..  
FT DOMAIN 23 499 EXTRACELLULAR (POTENTIAL).  
FT TRANSMEM 500 520 POTENTIAL.  
FT DOMAIN 521 555 CYTOPLASMIC (POTENTIAL).  
FT DOMAIN 31 127 IG-LIKE V-TYPE 1.  
FT DOMAIN 131 229 IG-LIKE V-TYPE 2.  
FT DOMAIN 323 397 IG-LIKE C2-TYPE 1.  
FT DOMAIN 319 397 IG-LIKE C2-TYPE 2.  
FT DOMAIN 406 484 IG-LIKE C2-TYPE 3.  
FT DISULFID 38 110 POTENTIAL.  
FT DISULFID 154 217 POTENTIAL.  
FT DISULFID 263 306 POTENTIAL.  
FT DISULFID 426 470 POTENTIAL.  
FT CARBOHYD 92 92 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 171 171 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 350 350 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 441 441 N-LINKED (GLCNAC. . .) (POTENTIAL).  
FT CARBOHYD 465 465 N-LINKED (GLCNAC. . .) (POTENTIAL).  
SQ SEQUENCE 555 AA; 60371 MW; 54A8014F00BF68 CRC64;

Query Match 4.2%; Score 144.5; DB 1; Length 555;  
Beet Local Similarity 19.1%; Pred. No. 0.058;  
Matches 118; Conservative 91; Mismatches 203; Indels 207; Gaps 30;

OY VLGGKGDVVELTCTASOKS---IQFHKNQNOIKIIGNOSFLLTKGSK-----LNDK 79  
DB VIGLYGETIVPCNDGTCKPDLFTTKK---YKDDSPDGLVKAQKDEATVATDG 82  
OY ADSRRSLWDQGNFPLIKLKIEDSDTYICEVEDQKEEVQLVFGLTANSDTHLQGS 139  
DB YKSNVSI--ANSSLLIARGLADQVFTC-----MVVSFTN-----LEEYSV 123  
OY 140 TLTLSPGSGSPVQCNPRKNIQGGKTLISVSOLELQDSTWTCTVLQNGKVEFKIDI 199  
DB 124 EVKHKRP-SAPVIK--NNAKELNGKLTQLGECVENA----- 159  
OY 200 VVLAFOKASSIVYKKEGQV-----EFSPLAFIVEKLTGSGELMWAERASSKSWITF 254  
DB 160 -----NPPADLIWKKNQTLVDDGKTIITSTTKDKTG-----LSTSSRLQY 204  
OY 255 DLKKNKSVKRVTDQPKLQNGKLLPLHLTLPQAVAGSNLTLEAKTKGKLLH---OE 311  
DB 205 TAKKEDVE-SQFTCAKIVMG---PDQVSEPSFP-----IHYETEK 242  
OY 312 VNLVVMRATQLOK---NLTCGVWGPTSPKMLSLKLENKAKVSKKEKPVVVL---NP 363  
DB 243 VSLQVVSQSPIREGEDVTLKCADGNPP--TSFENIKKQKVTATVDKDYTLTGVTBA 299  
OY 364 EAGWQCLLSDSGOVLLES-----NIVLTLTMSIPVPCAPRPSKCDKHTTCFELL 414  
DB 300 DSGIYKCSLLDND--WVESTQFTVVSFLDVSFLTGTG----- 334  
OY 415 GGPSPVFLPPKPKDTLMISRTPEVTCVVDVSHEDPEVKFMVYDVGVVHNAKTKPREED 474

DB 335 -----VLKNVGENLTVSLD-KNASSBAKVTWTKD-----NRKLDKLPDF 372  
OY 475 YNSTRYRVSVLTVLHQDWLNGKEKCKVKSKALPAPEKTIKSKAQGPAPROYTLPPSR 534  
DB 373 SKLTVSPAGL-----YCDVSTIEG-----IKRSLSELTVEGIPKITSLTGKR 415  
OY 535 DELTKQVSLTCLVKGFPSPDIAMWESNQPE-----NNYKTPPVLDSDGSPFLYSKLT 590  
DB 416 SSGDKHKV-LTCEAEBSPKRD--VQMSVNTNDEVSYNNKGAT-----YKLT 459  
OY 591 VDKSRMOQGNVFGSGVMHE 609  
DB 460 VVPSK--NLTVSCLVLTNK 475

RESULT 94  
SHS1 HUMAN STANDARD; PRT; 503 AA.  
AC P78324; O00683; O43799; Q8N517; Q8TAL8; Q9H022; Q9UDX2; Q9UIJ6;  
AC O9Y4U9;  
DT 10-OCT-2003 (Rel. 42, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DT 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Protein-tyrosine phosphatase non-receptor type substrate 1 precursor  
DE (SHP substrate-1) (SHPs-1) (inhibitory receptor SHP-1) (Signal-  
DE regulatory protein alpha-1) (SHP-alpha-1) (SHP-alpha-2) (SHP-alpha-  
DE 3) (MYD-1 antigen) (Brain Ig-like molecule with tyrosine-based  
DE activation motifs) (Blt) (Macrophage fusion receptor) (p84).  
GN PTPN1 OR SHS1 OR SIRP OR MYD1 OR BIT OR MFR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
CC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Brain;  
RX MEDLINE=97223399; PubMed=9070220;  
RA Ymiao T., Matozaki T., Amano K., Matsuda Y., Takahashi N., Ochi F.,  
RT Fujioka Y., Kasuga M.;  
RT "Mouse and human SHP-1: molecular cloning of cDNAs and chromosomal  
RT localization of genes.";  
RL Biochem. Biophys. Res. Commun. 231:61-67(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 1), POLYMORPHISM, PHOSPHORYLATION,  
RP N-GLYCOSYLATION, AND INTERACTIONS WITH PTPN11; PTPN6 AND GRB2.  
RC TISSUE=Placenta;  
RX MEDLINE=97215901; PubMed=9062191;  
RA Kharitonov A., Chen Z., Sures I., Wang H., Schilling J.,  
RT Ulrich A.;  
RT "A family of proteins that inhibit signalling through tyrosine kinase  
RT receptors.";  
RL Nature 386:181-186(1997).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 3), AND VARIANTS ASP-95; LEU-96; ASN-100;  
RP ARG-107; GLY-109 AND VAL-131.  
RC TISSUE=Monocytes;  
RX MEDLINE=98143722; PubMed=9485180;  
RA Brooke G.P., Parsons K.R., Howard C.J.;  
RT "Cloning of two members of the SIRP alpha family of protein tyrosine  
RT phosphatase binding proteins in cattle that are expressed on monocytes  
RT and a subpopulation of dendritic cells and which mediate binding to  
RT CD4 T cells.";  
RL Eur. J. Immunol. 28:1-11(1998).  
RN [4]  
RP SEQUENCE FROM N.A. (ISOFORM 1), AND VARIANTS LEU-44; THR-50; THR-52;  
RP ARG-54; ALA-57; GLY-75; ASP-95; LEU-96; ASN-100; ARG-107; GLY-109 AND  
RP VAL-131.  
RC TISSUE=Brain;  
RX MEDLINE=20053880; PubMed=10585853;  
RA Sano S.-I., Ohnishi H., Kubota M.;  
RT "Gene structure of mouse BIR/SHPS-1.";  
RL Biochem. J. 344:667-675(1999).  
RN [5]

RP SEQUENCE FROM N.A.  
 RX MEDLINE=1638749; PubMed=11780052;  
 RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
 RA Jones M., Stavrides G., Almeida A.K., Babbage A.K., Bagduly C.L.,  
 RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
 RA Bailely O.P., Bird C.P., Blakey S.E., Bridgman M., Brown A.J.,  
 RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,  
 RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
 RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
 RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
 RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
 RA Graffam D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
 RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
 RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
 RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
 RA Levasilho M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
 RA Marsh V.L., Martin S.L., McCormachie L.J., McElroy K., McMurray A.A.,  
 RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
 RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
 RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,  
 RA Rice C.M., Ross M.T., Scott C.E., Sena H.K., Showgreen R., Sims S.,  
 RA Snice C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,  
 RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
 RA Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M.,  
 RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,  
 RA Wilmshurst L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
 RA Rogers J.;  
 RT "The DNA sequence and comparative analysis of human chromosome 20.";  
 RL Nature 414:865-871(2001).  
 [6]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANTS LEU-44; THR-50;  
 RP THR-52; ARG-54; ALA-57; GLY-75; ASP-95; LEU-96; ASN-100; ARG-107;  
 RP GLY-109 AND VAL-131.  
 RC TISSUE=Brain, Kidney, and Skin;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Straube R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Kraemer R.D., Collins F.S., Wagner F.S., Sherman C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Heien F.,  
 RA Diatchenko L., Marusha K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Schaefer T.E.,  
 RA Brownstein M.J., Uedin T.B., Toshiyuki S., Cantinici P., Prange C.,  
 RA Raba S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullaby S.J.,  
 RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.V., Hulyk S.W.,  
 RA Villalón D.K., Muzny D.M., Sodegryn E.J., Lu X., Gibbs R.A.,  
 RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
 RA Whiting R.W., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butcherfield Y.S.N., Krzywinski M.I., Skalska U., Smalins D.E.,  
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length  
 RT human and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 [7]  
 RP FUNCTION AND INTERACTIONS WITH FYB, SCAP2 AND PTK2B.  
 RX MEDLINE=99401000; PubMed=10469599;  
 RA Tjims J.F., Swanson K.D., Marie-Cardine A., Raab M., Rudd C.E.,  
 RA Schraven B., Neel B.G.;  
 RT "SHPS-1 is a scaffold for assembling distinct adhesion-regulated  
 RT multi-protein complexes in macrophages.";  
 RL Curr. Biol. 9:927-930(1999).  
 [8]  
 RP PHOSPHORYLATION BY JAK2, AND INTERACTIONS WITH PTPN11 AND JAK2.  
 RX MEDLINE=20428747; PubMed=10842184;  
 RA Scofega M.R., Argersinger L.S., Wang H., Ullrich A., Carter-Su C.;  
 RT "Negative regulation of growth hormone receptor/JAK2 signaling by  
 RT signal regulatory protein alpha.";  
 RL J. Biol. Chem. 275:28222-28229(2000).  
 [9]  
 RP FUNCTION AND INTERACTION WITH CD47.  
 RX MEDLINE=21400825; PubMed=11505994;

RA Latour S., Tanaka H., Demeure C., Mateo V., Rubio M., Brown E.J.,  
 RA Maliszewski C., Lindberg F.P., Oldenborg A., Ullrich A.,  
 RA Delapess G., Sarfati M.;  
 RT "Bidirectional negative regulation of human T and dendritic cells by  
 RT CD47 and its cognate receptor signal-regulator protein-alpha.";  
 RT down-regulation of IL-12 responsiveness and inhibition of dendritic  
 RT cell activation.";  
 RL J. Immunol. 167:2547-2554(2001).  
 CC -1- FUNCTION: Immunoglobulin-like cell surface receptor for CD47. Acts  
 CC as docking protein and induces translocation of PTPN6, PTPN11 and  
 CC other binding partners from the cytosol to the plasma membrane.  
 CC Supports adhesion of cerebellar neurons, neurite outgrowth and  
 CC glial cell attachment. May play a key role in intracellular  
 CC signaling during synaptogenesis and in synaptic function (By  
 CC similarity). Involved in the negative regulation of receptor  
 CC tyrosine kinase-coupled cellular responses induced by cell  
 CC adhesion, growth factors or insulin. Mediates negative regulation  
 CC of phagocytosis, mast cell activation and dendritic cell  
 CC activation. CD47 binding prevents maturation of immature dendritic  
 CC cells and inhibits cytokine production by mature dendritic cells.  
 CC -1- SUBUNIT: Binds PTPN11 when tyrosine-phosphorylated, except in  
 CC macrophages, where it primarily binds PTPN6. Binds GRB2 in vitro.  
 CC Binds FGR (By similarity). Binds JAK2 irrespective of its  
 CC phosphorylation status and forms a stable complex. Binds SCAP1  
 CC and/or SCAP2. The resulting complex recruits FYB. Binds PTK2B.  
 CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
 CC -1- ALTERNATIVE PRODUCTS: Type I membrane protein.  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;  
 CC IsoId=P78324-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=P78324-2; Sequence=VSP\_007030;  
 CC Note=No experimental confirmation available;  
 CC Name=3;  
 CC IsoId=P78324-3; Sequence=VSP\_007029;  
 CC Note=No experimental confirmation available;  
 CC -1- TISSUE SPECIFICITY: Ubiquitous. Highly expressed in brain.  
 CC Detected on myeloid cells, but not T cells. Detected at lower  
 CC levels in heart, placenta, lung, testis, ovary, colon, liver,  
 CC small intestine, prostate, spleen, kidney, skeletal muscle and  
 CC pancreas.  
 CC -1- PTM: N-glycosylated.  
 CC -1- PTM: Phosphorylated on tyrosine residues in response to  
 CC stimulation with EGF, growth hormone, insulin and PDGF.  
 CC Dephosphorylated by PTPN11.  
 CC -1- SIMILARITY: Contains 2 immunoglobulin-like V-type domains.  
 CC -1- SIMILARITY: Contains 1 immunoglobulin-like V-type domain.  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
 CC -----  
 DR EMBL; D86043; BAA12974.1; -;  
 DR EMBL; Y10375; CAA71403.1; -;  
 DR EMBL; Y11047; CAA71944.1; ALT INIT.  
 DR EMBL; AB023430; BAA87929.1; -;  
 DR EMBL; AC004832; AAF19260.1; -;  
 DR EMBL; AL034562; CAB38874.1; -;  
 DR EMBL; AL049634; CAB46662.1; -;  
 DR EMBL; AL117335; CAC12723.1; -;  
 DR EMBL; BC026692; AAH26692.1; -;  
 DR EMBL; BC033092; AAH33092.1; -;  
 DR EMBL; BC038510; AAH38510.1; -;  
 DR PIR; JCS287; JCS287.  
 DR Genew; HGNC; 9662; PTPN11.  
 DR MIM; 602461; -;  
 DR GO; GO:0005886; C:plasma membrane; TAS.  
 Query Match 4.2%; Score 143.5; DB 1; Length 503;

Best Local Similarity 21.8%; Pred. No. 0.059;  
Matches 87; Conservative 54; Mismatches 166; Indels 93; Gaps 17;

QY KEVSKRYTODPKQWKKLPLHJTLPLAL-----QVAGSGL-TLLEKTKLKHQEV 312  
DQ 32 BELVQIQDQSVSAAGSALHCTVSLIVGPQWRGNGPARELYNQKGF---- 87  
QY 313 NLVVRATQLOKNTCEVWGPTSPKMLSLKLEKKAQSKREKRVWLNPEAGMOCLL 372  
DQ 88 -----PRVTVSSESTRKEMNDPSISN-----ITPADACTYCYK 123  
QY 373 SDGQVLE-----SNIKVLPMTSTVPCCPAPEKSCDKTHCTPELGGSEVLPFPKP 426  
DQ 124 FRKSPDTEFKSGAGTSLVRAKPSAPV-----VSGPARATP--- 161  
QY 427 KDTLMISTREVTGVVVDVSHEDPEVKMNVVDGVEVNAATK--PRE-----QYNSTYR 480  
DQ 162 -----QHTVSFTESHGSPRD--ITLKMFKNENELSDFTQNDVPGEVSYSIHSTAK 213  
QY 481 VSVLTVLHODMLNGKEYKCKVSNKALPA-PIEKT--ISKAGQPREPOVYTLPPSRDEL 537  
DQ 214 VVLVREDHSGVI-----CEVAHVTLQGDPLRGITANLSEITIRVPPILEV-TQDPVPAE- 265  
QY 538 TKQVSLTCLVKGFPSPDIADAVESWESNGQPPENNYKTPPEVLDSGSEFLYSKLTVDKSRWQ 597  
DQ 266 --NQVNTVCQVRFKYPQRLQLTWLENGVNSRTETASTVTEKDKGTVMMSVLLVNSAHR 323  
QY 598 QGNVFSQSVHME---ALHNHTOKSLSLSPGLQDERTCAE 634  
DQ 324 DVKLTQCVHEHGGQPAVSKSHDLK-VSAHPKEQGSNTAAE 362

RESULT 95  
CAML\_MOUSE  
ID CAML\_MOUSE STANDARD; PRT; 1260 AA.  
AC P11627;  
DT 01-OCT-1989 (Rel. 12, Created)  
DT 01-OCT-1989 (Rel. 12, Last sequence update)  
DT 15-MAR-2004 (Rel. 43, Last annotation update)  
DE Neural cell adhesion molecule L1 precursor (N-CAM L1).  
GN L1CAM OR CAML1.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Euteria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TextID=10090;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Brain;  
RX MEDLINE=88318924; PubMed=3412448;  
RA Moos M., Tacke R., Scherer H., Teplow D., Frueh K., Schachner M.;  
RT "Neural adhesion molecule L1 as a member of the immunoglobulin  
superfamily with binding domains similar to fibronectin.";  
RL Nature 334:701-703(1988).  
CC -1- FUNCTION: Cell adhesion molecule with an important role in the  
development of the nervous system. Involved in neuron-neuron  
adhesion, neurite fasciculation, outgrowth of neurites, etc. Binds  
to axonin on neurons.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- SIMILARITY: Belongs to the immunoglobulin superfamily.  
CC -1- SIMILARITY: Contains 6 immunoglobulin-like C2-type domains.  
CC -1- SIMILARITY: Contains 5 fibronectin type III domains.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (see <http://www.ebi.ac.uk/announcements>  
or send an email to [license@ebi.ac.uk](mailto:license@ebi.ac.uk)).  
CC -----  
DR EMBL; X12875; CAA31368.1; -;  
PIR; S05479; S05479.

DR HSP; P20241; ICFB.  
DR MG; MGI:96721; L1cam.  
DR GO; GO:0007411; P:axon guidance; IMP.  
DR InterPro; IPR006957; FN\_III-like.  
DR InterPro; IPR003961; FN\_III.  
DR InterPro; IPR007110; IG\_III.  
DR InterPro; IPR003598; IG\_C2.  
DR Pfam; PF00047; Ig; 6.  
DR Pfam; PF00047; Ig; 6.  
DR SMART; SM00060; FN3; 2.  
DR SMART; SM00408; IGc2; 5.  
DR PROSITE; PS50835; IG\_LIKE; 6.  
DR Neurogenesis; Cell adhesion; Developmental protein; Glycoprotein;  
KW Transmembrane; Repeat; Immunoglobulin domain; Signal.  
FT SIGNAL 1  
FT CHAIN 20 1260  
FT DOMAIN 20 1123  
FT TRANSMEM 1124 1146  
FT DOMAIN 1147 1260  
FT DOMAIN 35 130  
FT DOMAIN 138 225  
FT DOMAIN 239 327  
FT DOMAIN 332 419  
FT DOMAIN 424 506  
FT DOMAIN 517 600  
FT DOMAIN 827 896  
FT DOMAIN 932 1094  
FT DOMAIN 1032 1094  
FT SITE 553 555  
FT SITE 562 564  
FT DISULFID 57 113  
FT DISULFID 157 208  
FT DISULFID 263 311  
FT DISULFID 353 403  
FT DISULFID 447 496  
FT DISULFID 538 590  
FT CARBOHYD 100 100  
FT CARBOHYD 202 202  
FT CARBOHYD 246 246  
FT CARBOHYD 293 293  
FT CARBOHYD 432 432  
FT CARBOHYD 478 478  
FT CARBOHYD 489 489  
FT CARBOHYD 504 504  
FT CARBOHYD 587 587  
FT CARBOHYD 670 670  
FT CARBOHYD 725 725  
FT CARBOHYD 776 776  
FT CARBOHYD 824 824  
FT CARBOHYD 848 848  
FT CARBOHYD 875 875  
FT CARBOHYD 968 968  
FT CARBOHYD 978 978  
FT CARBOHYD 1022 1022  
FT CARBOHYD 1030 1030  
FT CARBOHYD 1073 1073  
FT CARBOHYD 1107 1107  
SQ SEQUENCE 1260 AA; 140968 MW; 22BE57001CB2A538 CRC64;  
Query Match 4.2%; Score 143.5; DB 1; Length 1260;  
Best Local Similarity 18.9%; Pred. No. 0.2;  
Matches 122; Conservative 94; Mismatches 252; Indels 179; Gaps 29;

QY 8 RHLLV---LQALPA-ATQGNRVLGKGDVETLTCTASQKSIQFHWKNSQIKI 61  
DQ 408 OHGILLANAVIYVQPARILTKDNTYMAVESSTAYLLCKARGAVPSVQWMDDEGTV 467  
QY 62 LQNGSELTGSPKANDRADRSRLNDQGFPLIKLTKEDSDTYTICEVEDKKEVQL 121  
DQ 468 LQDERFP-----PYANGTISIDLDLQNDYGRVFCQANDQNNVTLL 508  
QY 122 VFGLTANSDTHLQ-----QSLTTLTLESPGSSPSVQCR-----SPRGKN-- 162

D5	509	A-NLOUYEANQIYQGRSNAIEKKGAARVFTQASPDPSLQASITMRGGGRDLQDRGSDK	567
OY	163	--IQGKTLTSVSOLELQDSTGWTCTVLQNOQKVEFKIDIVLA-----FOKASITVYK	213
DB	568	YFIEDGK-LVIOQLDPSDQSNVSCVASTELDEVEERADLLVVGSGPVPNHELSDRHLLK	626
OY	214	KEGQVQFSS-----PFLAFVVEKLTQSGSLMMQAEBAASSKSMITFDLKNKEVS	262
DB	627	QSOVHLSWSAEEDHNSPIEKYDIEEFDEKEM--APEKMSLQKVNQNGQSTTLKLSPVNH	684
OY	263	VKRVTDPPKLOMQRKLP--HLTLPOALPQ-----YAASGNLTLLAEKTKLQOEVLV	315
DB	685	TFPRTAINTKGGPEPSPVSSVTPREAAPEKRPVDVREGN-----ETNNM	720
OY	316	VMRATQLOKMLTCEVGR-----TSPKMLMSLKLENKEAKVSKREKVPWVLNPEAQMQ	369
DB	731	VI-----TWKPLRLMMDMNAPOIQYVWQVRPOGQKBTWRKQTV-----	767
OY	370	CLASDSQVLLSESNIKVLPWSPVPC-----PAPERPSG-----DKHTHCEPLGGPS	418
DB	768	-----SPPFLVSNSTSTFPVPEYELKQVAVNNQSGPEPQVTTGYSGEDVPVSPFL----	817
OY	419	VFLPFPKPKD-TLMSIRTEPEVTCVVDVSHEDPEVK-FN---WYVDGVEVNAATKREE	473
DB	818	-----EDITFNSSTVLVRRRPDLQAVQGHKLGNNVYWMGSGRKNKSKHNKSH	866
OY	474	---QYNSTRVSVSVLTVLQDWMLNGKEYKCV--SNKALPAPLEKTIKSAKGQPREPV	527
DB	870	IIVPANTTASITLSGLRPY-----SSYHAEVQAFNGGGLGSPASWTFSTPEGVPHPEA	922
OY	528	YTLPRSPD-----ELTKQVSLTCLVKKFYSDDAVEWESNQ	565
DB	923	LHLECGSDTSLLLHMOPRLSHNGV-LTGYLLSYNP---VEGESKEQ	964
RESULT 96			
KAC_MOUSE			
ID	P01837	STANDARD;	PRT; 106 AA.
AC	21-JUL-1986	(Rel. 01, Created)	
DT	21-JUL-1986	(Rel. 01, Last sequence update)	
DT	10-OCT-2003	(Rel. 42, Last annotation update)	
DE	Ig kappa chain C region.		
OS	Mus musculus (Mouse).		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.		
OX	NCBI_TaxID=10090;		
RN			
RP	SEQUENCE (MOPC 21).		
RX	MEDLINE=73053310;	PubMed=4638343;	
RA	Svaesti J., Mlalestin C.;		
RT	"The complete amino acid sequence of a mouse kappa light chain.";		
RL	Biochem. J. 128:427-444(1972).		
RN			
RP	DISULFIDE BONDS (MOPC 21).		
RX	MEDLINE=73008889;	PubMed=5073237;	
RA	Svaesti J., Mlalestin C.;		
RT	"The disulphide bridges of a mouse immunoglobulin G1 protein.";		
RL	Biochem. J. 126:837-850(1972).		
RN			
RP	SEQUENCE FROM N.A., AND REVISIONS TO 53-59.		
RX	MEDLINE=79084137;	PubMed=103625;	
RA	Hamlyn P.H., Brownlee G.G., Cheng C.-C., Galt M.J., Mlalestin C.;		
RT	"Complete sequence of constant and 3' noncoding regions of an		
RT	immunoglobulin mRNA using the dideoxynucleotide method of RNA		
RT	sequencing.";		
RL	Cell 15:1067-1075(1978).		
RN			
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=82059477;	PubMed=6170937;	
RA	Hamlyn P.H., Galt M.J., Mlalestin C.;		
RT	"Complete sequence of an immunoglobulin mRNA using specific priming		
RT	and the dideoxynucleotide method of RNA sequencing.";		

```

RN Nucleic Acids Res. 9:4485-4494(1981).
RL (5)
RP SEQUENCE FROM N.A.
RA MEDLINE=81191915; PubMed=6262318;
RA Max E.B., Matzel J.V. Jr., Leder P.;
RT "The nucleotide sequence of a 5.5-kilobase DNA segment containing the
RT mouse kappa immunoglobulin J and C region genes.";
RL J. Biol. Chem. 256:5116-5120(1981).
RN (6)
RN SEQUENCE FROM N.A.
RX MEDLINE=81198949; PubMed=6785724;
RA Alenburger W., Neumaier P.S., Steimez M., Zachau H.G.;
RT "DNA sequence of the constant gene region of the mouse immunoglobulin
RT kappa chain.";
RL Nucleic Acids Res. 9:971-981(1981).
RN (7)
RN SEQUENCE FROM N.A.
RP MEDLINE=88329081; PubMed=3138116;
RA de Waele P., Feys V., van de Voorde A., Motemans F., Fiers W.;
RT "Expression in non-lymphoid cells of mouse recombinant immunoglobulin
RT directed against the tumour marker human placental alkaline
RT phosphatase.";
RL Eur. J. Biochem. 176:287-295(1988).
CC -I- SIMILARITY: Contains 1 immunoglobulin-like domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.ebi.ac.uk/announcement/
CC or send an email to license@ebi.ac.uk).
CC -----
CC EMBL: V00807; CAA24189.1; -.
DR PIR: B90262; KIMS.
DR PDB: 1A1F; 01-FEB-97.
DR PDB: 1ESK; 02-OCT-00.
DR PDB: 1K85; 08-APR-98.
DR PDB: 1KCR; 11-MAY-02.
DR PDB: 1KCU; 11-MAY-02.
DR PDB: 1KCV; 11-MAY-02.
DR PDB: 25C8; 09-JUL-99.
DR MGD: MGI:96495; I9K-C.
DR InterPro: IPR007110; IG-like.
DR InterPro: IPR003597; IG_c1.
DR InterPro: IPR003006; IG_MHC.
DR Pfam: PF00047; Ig; 1.
DR SMART: SM00407; TcG1; 1.
DR PROSITE: PS50835; IG_LIKE; 1.
DR PROSITE: PS00290; IG_MHC; 1.
KW Immunoglobulin domain; Immunoglobulin C region; 3D-structure.
FT FT DOMAIN 1 102 IG-LIKE.
FT DISULFID 26 86
FT DISULFID 106 106 INTERCHAIN (WITH A HEAVY CHAIN).
SQ SEQUENCE 106 AA; 11778 MW; 4851FF5E949BAEB5 CRC64;
Query Match 4.2%; Score 143; DB 1; length 106;
Beet local similarity 31.5%; Pred. No. 0.0083;
Matches 29; Conservative 18; Mismatches 33; Indels 12; Gaps 2;
QY 525 PQYTLPEPSRDELTKNOVSLTCLVKGFPYPSDIAVWESNGOPENNYKTPVLDLSDS----- 579
DB 5 PTVSIFPPSSSSQTLTKSGASVVCPLINFFPKDINVKMKDKDGERQN-----GVLSNWTDQD 59
QY 580 --DGSFFLYSLKLTVDKSRMQGQGNVFGSGVWHE 609
DB 60 SKDSIVSWSSTLTLLKDEYERHNSYTCRATH 91

```

ID SREB1 HUMAN STANDARD; PRT; 398 AA.  
AC O00241; O8TB12; O9HI05; O9Y4V0;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 10-OCT-2003 (Rel. 42, Last sequence update)  
DE 10-OCT-2003 (Rel. 42, Last annotation update)  
DE Signal-regulatory protein beta-1 precursor (SIRP-beta-1).  
GN SIRPB1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM 1).  
RC TISSUE=Placenta;  
RX MEDLINE=97215901; PubMed=9062191;  
RA Kharitonov A., Chen Z., Sures I., Wang H., Schilling J.,  
RT Ullrich A.;  
RT "A family of proteins that inhibit signalling through tyrosine kinase  
receptors.";  
RL Nature 386:181-186(1997).  
RN [2]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RX MEDLINE=21638749; PubMed=11780057;  
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,  
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,  
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,  
RA Beasley C.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,  
RA Buck D., Buttrill W.D., Butler A.P., Carder C., Carter N.P.,  
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,  
RA Cobley V.E., Collier R.E., Connor R.E., Corby N.R.,  
RA Coulson A., Coville G.J., Deadman R., Dhami P.D., Dunn M.,  
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,  
RA Grahm D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,  
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,  
RA Humble E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,  
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,  
RA Lehaeslath M.H., Leverisa M.A., Lloyd C., Lloyd D.M., Lovell J.D.,  
RA Marsh V.L., Martin S.L., McConachie L.J., McElay K., McMurtry A.A.,  
RA Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,  
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,  
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsey H.,  
RA Rice C.M., Rose M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,  
RA Stuce C.D., Smith M.L., Soderlund C., Stewart C.A., Sultun J.E.,  
RA Swann R.M., Symmore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,  
RA Tracey A., Tromans A.C., Vautin M., Wall M., Wallis J.M.,  
RA Whitehead S.L., Whitelaker P., Willey D.L., Williams L., Williams S.A.,  
RA Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,  
RT Rogers J.;  
RT "The DNA sequence and comparative analysis of human chromosome 20.";  
RL Nature 414:865-871(2001).  
RN [3]  
RP SEQUENCE FROM N.A. (ISOFORM 2).  
RC TISSUE=Skin;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
RA Krausberg R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant P.L., Scheetz T.E.,  
RA Brownstein M.J., Udell T.B., Tothiyuki S., Carninci P., Prange C.T.,  
RA Raha S.S., Loggellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,  
RA Bobak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
RA Richardson S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hultay S.W.,  
RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
RA Fahy J., Helton E., Kettman M., Madan A., Rodriguez S., Sanchez A.,  
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
RA Blakeley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.W.,  
RA Butterfield Y.S.N., Krzywinski M.I., Skalek A., Smillie D.E.,  
RA Schnerich A., Schein J.E., Jones S.J.M., Marra M.A.;  
RT "Generation and initial analysis of more than 15,000 full-length  
human and mouse cDNA sequences.";

RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
RN [4]  
RP INTERACTION WITH TYROBP AND SYK.  
RX PubMed=11169422;  
RA Tomasello E., Cant C., Buehring H.-J., Vely F., Andrie P., Seifert M.,  
RA Ullrich A., Vliet E.;  
RT "Association of signal-regulatory proteins beta with KAPAP/DAP-12.";  
RL Eur. J. Immunol. 30:2147-2156(2000).  
RN [5]  
RP INTERACTION WITH TYROBP, N-GLYCOSYLATION, AND TISSUE SPECIFICITY.  
RX MEDLINE=20072721; PubMed=10604985;  
RA Dietrich J., Cella M., Seifert M., Buehring H.-J., Colonna M.;  
RT "Signal-regulatory protein beta 1 is a DAP12-associated activating  
receptor expressed in myeloid cells.";  
RL J. Immunol. 164:9-12(2000).  
CC -1- FUNCTION: Immunoglobulin-like cell surface receptor involved in  
the negative regulation of receptor tyrosine kinase-coupled  
signaling processes. Participates also in the recruitment of  
tyrosine kinase SYK.  
CC -1- SUBUNIT: Interacts with TYROBP. This interaction results in the  
recruitment of SYK.  
CC -1- SUBCELLULAR LOCATION: Type I membrane protein.  
CC -1- ALTERNATIVE PRODUCTS:  
Name=1;  
Name=2;  
Name=3;  
Name=4;  
Name=5;  
Name=6;  
Name=7;  
Name=8;  
Name=9;  
Name=10;  
Name=11;  
Name=12;  
Name=13;  
Name=14;  
Name=15;  
Name=16;  
Name=17;  
Name=18;  
Name=19;  
Name=20;  
Name=21;  
Name=22;  
Name=23;  
Name=24;  
Name=25;  
Name=26;  
Name=27;  
Name=28;  
Name=29;  
Name=30;  
Name=31;  
Name=32;  
Name=33;  
Name=34;  
Name=35;  
Name=36;  
Name=37;  
Name=38;  
Name=39;  
Name=40;  
Name=41;  
Name=42;  
Name=43;  
Name=44;  
Name=45;  
Name=46;  
Name=47;  
Name=48;  
Name=49;  
Name=50;  
Name=51;  
Name=52;  
Name=53;  
Name=54;  
Name=55;  
Name=56;  
Name=57;  
Name=58;  
Name=59;  
Name=60;  
Name=61;  
Name=62;  
Name=63;  
Name=64;  
Name=65;  
Name=66;  
Name=67;  
Name=68;  
Name=69;  
Name=70;  
Name=71;  
Name=72;  
Name=73;  
Name=74;  
Name=75;  
Name=76;  
Name=77;  
Name=78;  
Name=79;  
Name=80;  
Name=81;  
Name=82;  
Name=83;  
Name=84;  
Name=85;  
Name=86;  
Name=87;  
Name=88;  
Name=89;  
Name=90;  
Name=91;  
Name=92;  
Name=93;  
Name=94;  
Name=95;  
Name=96;  
Name=97;  
Name=98;  
Name=99;  
Name=100;  
Name=101;  
Name=102;  
Name=103;  
Name=104;  
Name=105;  
Name=106;  
Name=107;  
Name=108;  
Name=109;  
Name=110;  
Name=111;  
Name=112;  
Name=113;  
Name=114;  
Name=115;  
Name=116;  
Name=117;  
Name=118;  
Name=119;  
Name=120;  
Name=121;  
Name=122;  
Name=123;  
Name=124;  
Name=125;  
Name=126;  
Name=127;  
Name=128;  
Name=129;  
Name=130;  
Name=131;  
Name=132;  
Name=133;  
Name=134;  
Name=135;  
Name=136;  
Name=137;  
Name=138;  
Name=139;  
Name=140;  
Name=141;  
Name=142;  
Name=143;  
Name=144;  
Name=145;  
Name=146;  
Name=147;  
Name=148;  
Name=149;  
Name=150;  
Name=151;  
Name=152;  
Name=153;  
Name=154;  
Name=155;  
Name=156;  
Name=157;  
Name=158;  
Name=159;  
Name=160;  
Name=161;  
Name=162;  
Name=163;  
Name=164;  
Name=165;  
Name=166;  
Name=167;  
Name=168;  
Name=169;  
Name=170;  
Name=171;  
Name=172;  
Name=173;  
Name=174;  
Name=175;  
Name=176;  
Name=177;  
Name=178;  
Name=179;  
Name=180;  
Name=181;  
Name=182;  
Name=183;  
Name=184;  
Name=185;  
Name=186;  
Name=187;  
Name=188;  
Name=189;  
Name=190;  
Name=191;  
Name=192;  
Name=193;  
Name=194;  
Name=195;  
Name=196;  
Name=197;  
Name=198;  
Name=199;  
Name=200;  
Name=201;  
Name=202;  
Name=203;  
Name=204;  
Name=205;  
Name=206;  
Name=207;  
Name=208;  
Name=209;  
Name=210;  
Name=211;  
Name=212;  
Name=213;  
Name=214;  
Name=215;  
Name=216;  
Name=217;  
Name=218;  
Name=219;  
Name=220;  
Name=221;  
Name=222;  
Name=223;  
Name=224;  
Name=225;  
Name=226;  
Name=227;  
Name=228;  
Name=229;  
Name=230;  
Name=231;  
Name=232;  
Name=233;  
Name=234;  
Name=235;  
Name=236;  
Name=237;  
Name=238;  
Name=239;  
Name=240;  
Name=241;  
Name=242;  
Name=243;  
Name=244;  
Name=245;  
Name=246;  
Name=247;  
Name=248;  
Name=249;  
Name=250;  
Name=251;  
Name=252;  
Name=253;  
Name=254;  
Name=255;  
Name=256;  
Name=257;  
Name=258;  
Name=259;  
Name=260;  
Name=261;  
Name=262;  
Name=263;  
Name=264;  
Name=265;  
Name=266;  
Name=267;  
Name=268;  
Name=269;  
Name=270;  
Name=271;  
Name=272;  
Name=273;  
Name=274;  
Name=275;  
Name=276;  
Name=277;  
Name=278;  
Name=279;  
Name=280;  
Name=281;  
Name=282;  
Name=283;  
Name=284;  
Name=285;  
Name=286;  
Name=287;  
Name=288;  
Name=289;  
Name=290;  
Name=291;  
Name=292;  
Name=293;  
Name=294;  
Name=295;  
Name=296;  
Name=297;  
Name=298;  
Name=299;  
Name=300;  
Name=301;  
Name=302;  
Name=303;  
Name=304;  
Name=305;  
Name=306;  
Name=307;  
Name=308;  
Name=309;  
Name=310;  
Name=311;  
Name=312;  
Name=313;  
Name=314;  
Name=315;  
Name=316;  
Name=317;  
Name=318;  
Name=319;  
Name=320;  
Name=321;  
Name=322;  
Name=323;  
Name=324;  
Name=325;  
Name=326;  
Name=327;  
Name=328;  
Name=329;  
Name=330;  
Name=331;  
Name=332;  
Name=333;  
Name=334;  
Name=335;  
Name=336;  
Name=337;  
Name=338;  
Name=339;  
Name=340;  
Name=341;  
Name=342;  
Name=343;  
Name=344;  
Name=345;  
Name=346;  
Name=347;  
Name=348;  
Name=349;  
Name=350;  
Name=351;  
Name=352;  
Name=353;  
Name=354;  
Name=355;  
Name=356;  
Name=357;  
Name=358;  
Name=359;  
Name=360;  
Name=361;  
Name=362;  
Name=363;  
Name=364;  
Name=365;  
Name=366;  
Name=367;  
Name=368;  
Name=369;  
Name=370;  
Name=371;  
Name=372;  
Name=373;  
Name=374;  
Name=375;  
Name=376;  
Name=377;  
Name=378;  
Name=379;  
Name=380;  
Name=381;  
Name=382;  
Name=383;  
Name=384;  
Name=385;  
Name=386;  
Name=387;  
Name=388;  
Name=389;  
Name=390;  
Name=391;  
Name=392;  
Name=393;  
Name=394;  
Name=395;  
Name=396;  
Name=397;  
Name=398;  
Name=399;  
Name=400;  
Name=401;  
Name=402;  
Name=403;  
Name=404;  
Name=405;  
Name=406;  
Name=407;  
Name=408;  
Name=409;  
Name=410;  
Name=411;  
Name=412;  
Name=413;  
Name=414;  
Name=415;  
Name=416;  
Name=417;  
Name=418;  
Name=419;  
Name=420;  
Name=421;  
Name=422;  
Name=423;  
Name=424;  
Name=425;  
Name=426;  
Name=427;  
Name=428;  
Name=429;  
Name=430;  
Name=431;  
Name=432;  
Name=433;  
Name=434;  
Name=435;  
Name=436;  
Name=437;  
Name=438;  
Name=439;  
Name=440;  
Name=441;  
Name=442;  
Name=443;  
Name=444;  
Name=445;  
Name=446;  
Name=447;  
Name=448;  
Name=449;  
Name=450;  
Name=451;  
Name=452;  
Name=453;  
Name=454;  
Name=455;  
Name=456;  
Name=457;  
Name=458;  
Name=459;  
Name=460;  
Name=461;  
Name=462;  
Name=463;  
Name=464;  
Name=465;  
Name=466;  
Name=467;  
Name=468;  
Name=469;  
Name=470;  
Name=471;  
Name=472;  
Name=473;  
Name=474;  
Name=475;  
Name=476;  
Name=477;  
Name=478;  
Name=479;  
Name=480;  
Name=481;  
Name=482;  
Name=483;  
Name=484;  
Name=485;  
Name=486;  
Name=487;  
Name=488;  
Name=489;  
Name=490;  
Name=491;  
Name=492;  
Name=493;  
Name=494;  
Name=495;  
Name=496;  
Name=497;  
Name=498;  
Name=499;  
Name=500;  
Name=501;  
Name=502;  
Name=503;  
Name=504;  
Name=505;  
Name=506;  
Name=507;  
Name=508;  
Name=509;  
Name=510;  
Name=511;  
Name=512;  
Name=513;  
Name=514;  
Name=515;  
Name=516;  
Name=517;  
Name=518;  
Name=519;  
Name=520;  
Name=521;  
Name=522;  
Name=523;  
Name=524;  
Name=525;  
Name=526;  
Name=527;  
Name=528;  
Name=529;  
Name=530;  
Name=531;  
Name=532;  
Name=533;  
Name=534;  
Name=535;  
Name=536;  
Name=537;  
Name=538;  
Name=539;  
Name=540;  
Name=541;  
Name=542;  
Name=543;  
Name=544;  
Name=545;  
Name=546;  
Name=547;  
Name=548;  
Name=549;  
Name=550;  
Name=551;  
Name=552;  
Name=553;  
Name=554;  
Name=555;  
Name=556;  
Name=557;  
Name=558;  
Name=559;  
Name=560;  
Name=561;  
Name=562;  
Name=563;  
Name=564;  
Name=565;  
Name=566;  
Name=567;  
Name=568;  
Name=569;  
Name=570;  
Name=571;  
Name=572;  
Name=573;  
Name=574;  
Name=575;  
Name=576;  
Name=577;  
Name=578;  
Name=579;  
Name=580;  
Name=581;  
Name=582;  
Name=583;  
Name=584;  
Name=585;  
Name=586;  
Name=587;  
Name=588;  
Name=589;  
Name=590;  
Name=591;  
Name=592;  
Name=593;  
Name=594;  
Name=595;  
Name=596;  
Name=597;  
Name=598;  
Name=599;  
Name=600;  
Name=601;  
Name=602;  
Name=603;  
Name=604;  
Name=605;  
Name=606;  
Name=607;  
Name=608;  
Name=609;  
Name=610;  
Name=611;  
Name=612;  
Name=613;  
Name=614;  
Name=615;  
Name=616;  
Name=617;  
Name=618;  
Name=619;  
Name=620;  
Name=621;  
Name=622;  
Name=623;  
Name=624;  
Name=625;  
Name=626;  
Name=627;  
Name=628;  
Name=629;  
Name=630;  
Name=631;  
Name=632;  
Name=633;  
Name=634;  
Name=635;  
Name=636;  
Name=637;  
Name=638;  
Name=639;  
Name=640;  
Name=641;  
Name=642;  
Name=643;  
Name=644;  
Name=645;  
Name=646;  
Name=647;  
Name=648;  
Name=649;  
Name=650;  
Name=651;  
Name=652;  
Name=653;  
Name=654;  
Name=655;  
Name=656;  
Name=657;  
Name=658;  
Name=659;  
Name=660;  
Name=661;  
Name=662;  
Name=663;  
Name=664;  
Name=665;  
Name=666;  
Name=667;  
Name=668;  
Name=669;  
Name=670;  
Name=671;  
Name=672;  
Name=673;  
Name=674;  
Name=675;  
Name=676;  
Name=677;  
Name=678;  
Name=679;  
Name=680;  
Name=681;  
Name=682;  
Name=683;  
Name=684;  
Name=685;  
Name=686;  
Name=687;  
Name=688;  
Name=689;  
Name=690;  
Name=691;  
Name=692;  
Name=693;  
Name=694;  
Name=695;  
Name=696;  
Name=697;  
Name=698;  
Name=699;  
Name=700;  
Name=701;  
Name=702;  
Name=703;  
Name=704;  
Name=705;  
Name=706;  
Name=707;  
Name=708;  
Name=709;  
Name=710;  
Name=711;  
Name=712;  
Name=713;  
Name=714;  
Name=715;  
Name=716;  
Name=717;  
Name=718;  
Name=719;  
Name=720;  
Name=721;  
Name=722;  
Name=723;  
Name=724;  
Name=725;  
Name=726;  
Name=727;  
Name=728;  
Name=729;  
Name=730;  
Name=731;  
Name=732;  
Name=733;  
Name=734;  
Name=735;  
Name=736;  
Name=737;  
Name=738;  
Name=739;  
Name=740;  
Name=741;  
Name=742;  
Name=743;  
Name=744;  
Name=745;  
Name=746;  
Name=747;  
Name=748;  
Name=749;  
Name=750;  
Name=751;  
Name=752;  
Name=753;  
Name=754;  
Name=755;  
Name=756;  
Name=757;  
Name=758;  
Name=759;  
Name=760;  
Name=761;  
Name=762;  
Name=763;  
Name=764;  
Name=765;  
Name=766;  
Name=767;  
Name=768;  
Name=769;  
Name=770;  
Name=771;  
Name=772;  
Name=773;  
Name=774;  
Name=775;  
Name=776;  
Name=777;  
Name=778;  
Name=779;  
Name=780;  
Name=781;  
Name=782;  
Name=783;  
Name=784;  
Name=785;  
Name=786;  
Name=787;  
Name=788;  
Name=789;  
Name=790;  
Name=791;  
Name=792;  
Name=793;  
Name=794;  
Name=795;  
Name=796;  
Name=797;  
Name=798;  
Name=799;  
Name=800;  
Name=801;  
Name=802;  
Name=803;  
Name=804;  
Name=805;  
Name=806;  
Name=807;  
Name=808;  
Name=809;  
Name=810;  
Name=811;  
Name=812;  
Name=813;  
Name=814;  
Name=815;  
Name=816;  
Name=817;  
Name=818;  
Name=819;  
Name=820;  
Name=821;  
Name=822;  
Name=823;  
Name=824;  
Name=825;  
Name=826;  
Name=827;  
Name=828;  
Name=829;  
Name=830;  
Name=831;  
Name=832;  
Name=833;  
Name=834;  
Name=835;  
Name=836;  
Name=837;  
Name=838;  
Name=839;  
Name=840;  
Name=841;  
Name=842;  
Name=843;  
Name=844;  
Name=845;  
Name=846;  
Name=847;  
Name=848;  
Name=849;  
Name=850;  
Name=851;  
Name=852;  
Name=853;  
Name=854;  
Name=855;  
Name=856;  
Name=857;  
Name=858;  
Name=859;  
Name=860;  
Name=861;  
Name=862;  
Name=863;  
Name=864;  
Name=865;  
Name=866;  
Name=867;  
Name=868;  
Name=869;  
Name=870;  
Name=871;  
Name=872;  
Name=873;  
Name=874;  
Name=875;  
Name=876;  
Name=877;  
Name=878;  
Name=879;  
Name=880;  
Name=881;  
Name=882;  
Name=883;  
Name=884;  
Name=885;  
Name=886;  
Name=887;  
Name=888;  
Name=889;  
Name=890;  
Name=891;  
Name=892;  
Name=893;  
Name=894;  
Name=895;  
Name=896;  
Name=897;  
Name=898;  
Name=899;  
Name=900;  
Name=901;  
Name=902;  
Name=903;  
Name=904;  
Name=905;  
Name=906;  
Name=907;  
Name=908;  
Name=909;  
Name=910;  
Name=911;  
Name=912;  
Name=913;  
Name=914;  
Name=915;  
Name=916;  
Name=917;  
Name=918;  
Name=919;  
Name=920;  
Name=921;  
Name=922;  
Name=923;  
Name=924;  
Name=925;  
Name=926;  
Name=927;  
Name=928;  
Name=929;  
Name=930;  
Name=931;  
Name=932;  
Name=933;  
Name=934;  
Name=935;  
Name=936;  
Name=937;  
Name=938;  
Name=939;  
Name=940;  
Name=941;  
Name=942;  
Name=943;  
Name=944;  
Name=945;  
Name=946;  
Name=947;  
Name=948;  
Name=949;  
Name=950;  
Name=951;  
Name=952;  
Name=953;  
Name=954;  
Name=955;  
Name=956;  
Name=957;  
Name=958;  
Name=959;  
Name=960;  
Name=961;  
Name=962;  
Name=963;  
Name=964;  
Name=965;  
Name=966;  
Name=967;  
Name=968;  
Name=969;  
Name=970;  
Name=971;  
Name=972;  
Name=973;  
Name=974;  
Name=975;  
Name=976;  
Name=977;  
Name=978;  
Name=979;  
Name=980;  
Name=981;  
Name=982;  
Name=983;  
Name=984;  
Name=985;  
Name=986;  
Name=987;  
Name=988;  
Name=989;  
Name=990;  
Name=991;  
Name=992;  
Name=993;  
Name=994;  
Name=995;  
Name=996;  
Name=997;  
Name=998;  
Name=999;  
Name=1000;  
Name=1001;  
Name=1002;  
Name=1003;  
Name=1004;  
Name=1005;  
Name=1006;  
Name=1007;  
Name=1008;  
Name=1009;  
Name=1010;  
Name=1011;  
Name=1012;  
Name=1013;  
Name=1014;  
Name=1015;  
Name=1016;  
Name=1017;  
Name=1018;  
Name=1019;  
Name=1020;  
Name=1021;  
Name=1022;  
Name=1023;  
Name=1024;  
Name=1025;  
Name=1026;  
Name=1027;  
Name=1028;  
Name=1029;  
Name=1030;  
Name=1031;  
Name=1032;  
Name=1033;  
Name=1034;  
Name=1035;  
Name=1036;  
Name=1037;  
Name=1038;  
Name=1039;  
Name=1040;  
Name=1041;  
Name=1042;  
Name=1043;  
Name=1044;  
Name=1045;  
Name=1046;  
Name=1047;  
Name=1048;  
Name=1049;  
Name=1050;  
Name=1051;  
Name=1052;  
Name=1053;  
Name=1054;  
Name=1055;  
Name=1056;  
Name=1057;  
Name=1058;  
Name=1059;  
Name=1060;  
Name=1061;  
Name=1062;  
Name=1063;  
Name=1064;  
Name=1065;  
Name=1066;  
Name=1067;  
Name=1068;  
Name=1069;  
Name=1070;  
Name=1071;  
Name=1072;  
Name=1073;  
Name=1074;  
Name=1075;  
Name=1076;  
Name=1077;  
Name=1078;  
Name=1079;  
Name=1080;  
Name=1081;  
Name=1082;  
Name=1083;  
Name=1084;  
Name=1085;  
Name=1086;  
Name=1087;  
Name=1088;  
Name=1089;  
Name=1090;  
Name=1091;  
Name=1092;  
Name=1093;  
Name=1094;  
Name=1095;  
Name=1096;  
Name=1097;  
Name=1098;  
Name=1099;  
Name=1100;  
Name=1101;  
Name=1102;  
Name=1103;  
Name=1104;  
Name=1105;  
Name=1106;  
Name=1107;  
Name=1108;  
Name=1109;  
Name=1110;  
Name=1111;  
Name=1112;  
Name=1113;  
Name=1114;  
Name=1115;



```

FT DOMAIN 234 326 IG-LIKE C2-TYPE 3.
FT DOMAIN 331 416 IG-LIKE C2-TYPE 4.
FT DOMAIN 426 522 FIBRONECTIN TYPE-III 1.
FT DOMAIN 525 618 FIBRONECTIN TYPE-III 2.
FT DOMAIN 619 716 FIBRONECTIN TYPE-III 3.
FT DOMAIN 722 816 FIBRONECTIN TYPE-III 4.
FT DOMAIN 840 940 FIBRONECTIN TYPE-III 5.
FT DOMAIN 941 1042 FIBRONECTIN TYPE-III 6.
FT DISULFID 61 117 BY SIMILARITY.
FT DISULFID 161 212 BY SIMILARITY.
FT DISULFID 261 310 BY SIMILARITY.
FT DISULFID 352 400 BY SIMILARITY.
FT CARBOHYD 94 94 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 299 299 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 318 318 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 478 478 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 628 628 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 702 702 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT VARIANT 168 168 M -> T (in oesophageal carcinoma).
FT VARIANT 201 201 R -> G.
FT VARIANT 201 201 /FTid=VAR_003910.
FT VARIANT 1375 1375 P -> H (in a colorectal carcinoma).
FT CONFLICT 138 138 /FTid=VAR_003911.
FT CONFLICT 233 329 MISSING (IN REF. 3).
FT CONFLICT 421 421 MISSING (IN REF. 3).
FT CONFLICT 421 421 MISSING (IN REF. 3).
SQ SEQUENCE 1447 AA; 158456 MM; 4A8612765ED0471F CRC64;

```

```

Query Match 4.1%; Score 141.5; DB 1; Length 1447;
Best Local Similarity 18.7%; Pred. No. 0.32;
Matches 131; Conservative 82; Mismatches 225; Indels 263; Gaps 33;

```

```

QY 34 GDTYELCTTSQKSIQFHWNKSNQIKILNQSGFLTKPSKLNDRADRSRLMDQGNP 93
DB 154 GDTYELCTTSQKSIQFHWNKSNQIKILNQSGFLTKPSKLNDRADRSRLMDQGNP 93
QY 94 LIINKLIEDSDTYCEVED-----QKEVOLL-----VFGLTANSDFHLQSGS 138
DB 197 LQISRLPQDGIYRCSARNPASRTGNEAVRLSLDPGLHRQLYFQRPNNVALLSGKO 256
QY 139 LTL---TLSEPPGS-----SPSVQCRSPRGKNIQSGKTLVSQLELDSGTWTCTVLON 189
DB 257 AVLECCSGYPRPFMTLRGEVYQLRSKK-YSLGGSNLLISVNTDDSGMTYCVTVTK 315
QY 190 QKKVEFKIDIVLA--FQKASIVYKKEGQVEFSPPLA----- 226
DB 316 NENISAGAEELTVLPWFLNHPNMLYVESMDIEFFECTVSGKPVPTVMMKNGDVI 375
QY 227 -----FTVEKLTGSGELMNG--AEKASSKSWITTFDLKXKEVSVKVTQPKQMG 275
DB 376 YFOVIGGSNRLRILGVKXSDGEFYQCAENENG-----NAQTSAGLLVPRKALPSS 425
QY 276 KKLPLHLTLPOAL-----PQYAGSGNL--TLAEATKGLQHEVNLVWM 317
DB 426 SVLP---SARDVVPVLVSSRFVRLSMRPRAEAKGNITQTFVFSREGDNRERLNTOP 482
QY 318 RAFOLO-KNLTCEV-----WGP-----TSPKMLSLKLENKEAKVSKRE 355
DB 483 GSIQLTGNLKEPAMTYFRVAVANEMGREGSSOPIKVATQELQVGPVERVLQA----- 536
QY 356 KPVVNLVPEAGMOCCLSDGQVLESINKLPTWSTPVPVPAEP---KSCDKTHRCPE 412
DB 537 -----VSTSPSTLITLWEPAYANGVQVRLPCTEVSQKE 573
QY 413 -----LIG-----GPSVFLFPKPKXTLMSRTPEVTCVV- 443
DB 574 QNIEVDGLSYLDEJLKKTEYSLARFLAVNRKGPV-----STDDITVTVTLIS 619
QY 444 DVSHEDPE-----VKFNW-----YNDGVEVNAKTRPRE----QYNST 478
DB 620 DVSPAPQONVSLLEVNSRSIKVSWLPPPSGTQNGFITGYKIRHAKTRRGEMETLEPNL 679

```

```

QY 479 YRVVSLTVLHOPMLNGKCKVSNKALPAPIETKISKAGOPREOVYTLPPSRDEL 538
DB 680 WYLFITGLE-----KSGVSPQVSAMTV-----NQTGPP--SNWTAETPENDLD 721
QY 539 KNOVSLTCLVKGFPSPDIKAVESNQPENN---YKTPPV 576
DB 722 ESQVP-----DQPSGLHV-----RPQNTCLIMSWTPL 749

RESULT 99
SNE2 HUMAN
ID SNE2 HUMAN STANDARD; PRT; 6885 AA.
AC Q8WXH0; Q8N1S3; Q8NF49; Q8TER7; Q8MWMW3; Q8MWMW4; Q8MWMW5; Q8WXH1;
AC Q8N1S0; Q8N1S4; Q8N1S5; Q8N1S6; Q8N1S7; Q8N1S8; Q8N1S9; Q8N1S10;
DT 10-OCT-2003 (Rel. 42, Created)
DT 10-OCT-2003 (Rel. 42, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Nesprin 2 (Nuclear envelope spectrin repeat protein 2) (Syne-2)
DE (Synaptic nuclear envelope protein 2) (Nucleus and actin connecting
DE element protein) (NUANCE protein).
CN SYNE2 OR NUA OR KIA1011.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A. (ISOFORMS 1 AND 9), FUNCTION, CHARACTERIZATION, AND
INTERACTION WITH F-ACTIN.
RX MEDLINE=22113122; PubMed=12118075;
RA Zhen Y.-Y., Libotte T., Munck M., Noegel A.A., Korenbaum E.;
RT "NUANCE, a giant protein connecting the nucleus and actin
RT cytoskeleton."
RL J. Cell Sci. 115:3207-3222(2002).
[2]
SEQUENCE FROM N.A. (ISOFORMS 4; 5 AND 7).
RX MEDLINE=21652858; PubMed=1172814;
RA Zhang Q., Skepper J.N., Yang F., Davies J.D., Hegyi L., Roberts R.G.,
RA Weissberg P.L., Ellis J.A., Shanahan C.M.;
RT "Nesprins: a novel family of spectrin-repeat-containing proteins that
RT localize to the nuclear membrane in multiple tissues."
RL J. Cell Sci. 114:4485-4498(2001).
[3]
SEQUENCE FROM N.A. (ISOFORM 1).
RX MEDLINE=22296983; PubMed=12408964;
RA Zhang Q., Regnault C., Greener M.J., Shanahan C.M., Roberts R.G.;
RT "The nesprins are giant actin-binding proteins, orthologous to
RT Drosophila melanogaster muscle protein MSP-300."
RL Genomics 80:473-481(2002).
[4]
SEQUENCE FROM N.A. (ISOFORM 6).
RX MEDLINE=12508121;
RA Helling R., Eckenberg R., Petit J.-L., Fonknechten N., Da Silva C.,
RA Cattoletto L., Levy M., Barbe V., De Bernardinis V., Ureta-Vidal A.,
RA Pelletier E., Vico V., Anthonard V., Rowen L., Madan A., Qin S.,
RA Sun H., Du H., Pepin K., Artigues F., Robert C., Grunau C.,
RA Bruehl T., Jallion O., Friedlander L., Samson G., Brotier P.,
RA Cure S., Segreus B., Aniere F., Samain S., Crespeau H., Abbasi N.,
RA Aach N., Boscus D., Dickhoff R., Dora W., Dubois I., Friedman C.,
RA Gouvenoux M., James R., Madan A., Malety-Estrada B., Mangenot C.,
RA Vacherie B., Bellemere C., Belser C., Bernard-Gonet M.,
RA Bitoi-Mavel D., Bourard M., Brizet-Silla S., Combette S.,
RA Dufosse-Laurent V., Perron C., Lechapis C., Louesse C., Muelet D.,
RA Magdelat G., Patteu B., Petit E., Stravain-Truklewicz P., Trybou A.,
RA Vega-Czarny N., Bataille E., Bluet E., Bortolais I., Dubois M.,
RA Dumont C., Guerin T., Haffray S., Hammedi R., Munga J., Pellouin V.,
RA Robert D., Wunderle E., Gauguet G., Roy A., Sainte-Marthe L.,
RA Verdier J., Verdier-Discala C., Hillier L.W., Fulton L., Mepherson J.,

```







```

OY 152 SVQCSPPGKNIQGGKTTISVSQLEL-----DQSGW--TCYTLQNOCK--VEFKXDI 199
DB 648 SKELR---LNKRMKRLVSKTLENNLPLMTKKQOPFDNSGNLISBEKRTVEFSTDW 704
OY 200 VLAFOKASSIV-----YKCEGOVEFSEFLPAFTYEKLTGSGELMWQABRASSKSWITP 254
DB 705 SVELPENYNNQNIKAGEKHEKENEFEPTGQKVAQVEKLIQGEI-MEAE---AKSVLDQ 759
OY 225 D-LNKNEVSXVRV-----TQPKLQMGKQLPLHLTPQALPOLYAGSGNLTALAEKTKL 308
DB 760 DDVDTSMEESLTHLAKGSMFDELARSBDMLQNDIONISSQESPOHYLTGLQAKIQA 819
OY 309 HOEVLVYMRATQLOKNTLCEVWGPTSPPLMTSLTENKEAKVSKREKPVWTLNPEAGMW 368
DB 820 KEKVQINVVKLIATLAKNLT-----DVSFDLDIRLMEF-----SQLELSYWMRAQQLIG 869
OY 369 QC-----LLSDSGOVLLESNIKVL 387
DB 870 QRESPGLISKKHEALISNTKSL 893

RESULT 100
HB2X_HUMAN STANDARD; PRT; 268 AA.
ID ID HB2X_HUMAN STANDARD; PRT; 268 AA.
AC P05538;
DT 01-NOV-1998 (Rel. 09, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE HLA class II histocompatibility antigen, DX beta chain precursor.
GN HLA-DQB2.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OC NCBI_TaxId=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=67250501; PubMed=3036828;
RA Jonsson A.K., Hyldig-Nielsen J.J., Servenius B., Larhammar D.,
RA Andersson G., Joergensen F., Peterson P.A., Raak L.;
RT "Class II genes of the human major histocompatibility complex.
RT Comparisons of the DQ and DX alpha and beta genes.";
RL J. Biol. Chem. 262:8767-8777(1987).
RN [2]
RP SEQUENCE OF 38-125 FROM N.A.
RX MEDLINE=85216510; PubMed=3558830;
RA Okada K., Boss J.W., Prentice H., Spies T., Mengler R., Auffray C.,
RA Little J.W., Grossberger D., Strominger J.L.;
RT "Gene organization of DC and DX subregions of the human major
RT histocompatibility complex.";
RL Proc. Natl. Acad. Sci. U.S.A. 82:3410-3414(1985).
CC CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@isb-sdb.ch).
CC -----
CC EMBL, M29614; -; NOT_ANNOTATED_CDS.
CC EMBL, M29615; -; NOT_ANNOTATED_CDS.
CC EMBL, M1136; -; NOT_ANNOTATED_CDS.
CC PIR, D29312; D29312.
CC HSSP, P13760; 2SBB.
CC GeneW, HGNC:4945; HLA-DQB2.
CC GO, GO:0016020; C:membrane; NAS.
CC GO, GO:0045012; F:MHC class II receptor activity; NAS.
CC GO, GO:0006955; P:immune response; NAS.
CC InterPro, IPR007110; Ig-like.
CC InterPro, IPR003597; Ig_C1.
CC InterPro, IPR003006; Ig_MHC.
CC InterPro, IPR000353; MHC II beta.

```

[illegible]

Search completed: August . 3, 2004, 13:09:24  
Job time : 14.8991 secs

```

QY      152 SVQCSPRGKNTGGKTLTSVSLQEL-----DPSGTM--TCTYLQNOCK--VEFKIDI 199
Db      648 SKELR---LNRKRWKLVSKTDLNNPLPMIKQDQPTFDNSGNILSBEKATVEFSTDW 704
QY      200 VLAFQKASSIV-----YKGEQVFEFSEPLAFTEKLTGSGELMWMQARASSSKGWTTF 254
Db      705 SVELPENNNQNTKAGEKHEKHEEFTTGOLKAKVDVEKILGQVEI-WEAE-----AKSVLDQ 759
QY      255 D-LKNKEVSKRV-----TQDPKLOMGKKLPLHLTPALPOLYAGSGNLTALAEKTKL 308
Db      760 DDVDTSMEESTLGHLLAKSGMPDELARASEDMQMDIQNISQSGESFOHYLTGLQAKIOEA 819
QY      309 HOEVVLVYMRATQLOKNTLCEVWGTPSPRLTSLTLENKEAVNSGRERPVWLTNEAGMW 368
Db      820 KEKVQINVAKLIALKLNLT-----DVSPLDILRLKME-----SQKELSYVMRAQQLLG 869
QY      369 QC-----LLSDSGQVLESNIKVTL 387
Db      870 QRESPELISKHEKALISNTKSL 893

RESULT 100
HB2X_HUMAN
ID      HB2X_HUMAN      STANDARD;      PRT;      268 AA.
AC      P05538;
DT      01-NOV-1988 (rel. 09, Created)
DT      01-FEB-1991 (rel. 17, Last sequence update)
DT      28-FEB-2003 (rel. 41, Last annotation update)
DE      HLA class II histocompatibility antigen, DX beta chain precursor.
GN      HLA-DQB2.
OS      Homo sapiens (Human).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX      NCBI_Taxid=9606;
[1]
RP      SEQUENCE FROM N.A.
RX      MEDLINE=87250501; Pubmed=3036828;
RA      Johnson A.K., Hyldig-Nielsen J.J., Serenius B., Larhammar D.,
RA      Andersson G., Joergensen F., Peterson P.A., Raak L.;
RT      "Class II genes of the human major histocompatibility complex.
RT      Comparisons of the DQ and DX alpha and beta genes.";
RT      J. Biol. Chem. 262:8767-8777(1987).
[2]
RN      SEQUENCE OF 38-125 FROM N.A.
RX      MEDLINE=85216510; Pubmed=3858830;
RA      Okada K., Bose J.M., Prentice H., Spies T., Mengler R., Auftray C.,
RA      Lilie U.W., Grosberger D., Strominger J.L.;
RT      "Gene organization of DC and DX subregions of the human major
RT      histocompatibility complex.";
RT      Proc. Natl. Acad. Sci. U.S.A. 82:3410-3414(1985).
[3]
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (see http://www.isb.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC
CC      -----
CC      EMBL, M29614; -, NOT_ANNOTATED_CDS.
CC      EMBL, M29615; -, NOT_ANNOTATED_CDS.
CC      EMBL, M11136; -, NOT_ANNOTATED_CDS.
CC      PIR, D29312; D29312.
CC      HSSP, P13760; 2SERB.
CC      Genew; HGNC:4945; HLA-DQB2.
CC      GO; GO:0016020; C:membrane; NAS.
CC      DR GO; GO:0045012; F:MHC class II receptor activity; NAS.
CC      DR GO; GO:0006955; P:immune response; NAS.
CC      DR InterPro; IPR007110; Ig-like.
CC      DR InterPro; IPR003597; Ig_C1.
CC      DR InterPro; IPR003006; Ig_MHC.
CC      DR InterPro; IPR000353; MHC II beta.

```

**THIS PAGE BLANK (uspro)**

GenCore version 5.1.6  
Copyright (c) 1993 - 2004 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 3, 2004, 13:00:24 ; Search time 14.8924 Seconds  
(without alignments)  
4185.504 Million cell updates/sec

Title: SEQ7  
Perfect score: 3414  
Sequence: 1 MNRGVPRHLLVLQLALLP.....DETCAEADGELDGLWTTDP 648

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 125 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2015	59.0	458	1 RWHUT4	T-cell surface gly
2	1867	54.7	432	1 RWCZT4	T-cell surface gly
3	1729	50.6	432	1 RWMQT4	T-cell surface gly
4	1211.5	35.5	330	1 GHU	Ig gamma-1 chain C
5	1205.5	35.3	374	2 S69339	Ig heavy chain V r
6	1203.5	35.3	255	4 S31865	Ig gamma-1 chain C
7	1171.5	34.3	234	2 PTO207	Ig gamma-1 chain C
8	1163	34.1	377	2 A23511	Ig gamma-3 chain C
9	1161	34.0	377	2 A60764	Ig gamma-3 chain C
10	1150.5	33.7	459	2 A46254	CD4 precursor - ra
11	1143	33.5	289	1 G3HUT1	Ig gamma-3 heavy c
12	1115.5	32.7	326	1 G2HU	Ig gamma-2 chain C
13	1105.5	32.4	327	1 G4HU	Ig gamma-4 chain C
14	1099	32.2	432	2 S30193	T-cell surface gly
15	1000.5	29.3	457	2 A27449	T-cell surface gly
16	993.5	29.1	457	1 RWMST4	T-cell surface gly
17	938.5	27.5	398	1 G3MSM	Ig gamma-3 chain C
18	928.5	27.2	393	1 G3MSM	Ig gamma-1 chain C
19	910	26.7	323	1 GHRB	Ig gamma-1 chain C
20	902.5	26.4	339	1 G2MSAM	Ig gamma-2a chain
21	901.5	26.4	328	2 I47160	Ig gamma-2b chain
22	901.5	26.4	328	2 I47159	Ig gamma-2a chain
23	894	26.2	277	2 I47162	Ig gamma-2 chain C
24	879.5	25.8	329	1 G2GP	Ig gamma-2 chain C
25	879.5	25.8	405	1 G2MSBM	Ig gamma-2b chain
26	876.5	25.7	328	2 I47161	Ig gamma-1 chain C
27	872.5	25.6	328	2 I47158	Ig gamma-1 chain C
28	870.5	25.5	470	2 S22080	Ig heavy chain pre
29	843	24.7	472	2 S31459	Ig gamma-1 chain -

30	840.5	24.6	329	1 G3MSC	Ig gamma-3 chain C
31	836	24.5	308	2 C30554	Ig heavy chain C r
32	830.5	24.3	444	2 PC4436	monoclonal antibody
33	829	24.3	326	2 PS0017	Ig gamma-1 chain C
34	826.5	24.2	333	2 PS0018	Ig gamma-2b chain
35	823.5	24.1	469	2 S37483	Ig gamma-2a chain
36	822.5	23.9	329	2 S00847	Ig gamma-2c chain
37	814.5	23.6	324	1 G2MS11	Ig gamma-1 chain C
38	807	23.5	474	1 G2MSA	Ig gamma-2a chain
39	801.5	23.5	330	1 G2MSA	Ig gamma-2a chain
40	798	23.4	446	2 S40295	Ig gamma-2a chain
41	792	23.2	322	2 PS0019	Ig gamma-2a chain
42	788	23.1	335	1 G2MSAB	Ig gamma-2a chain
43	779.5	22.8	475	2 S01321	Ig gamma-2b chain
44	768	22.5	327	2 S06611	Ig gamma-2 chain C
45	702	20.6	180	2 I46732	Ig gamma heavy cha
46	574.5	16.8	218	2 A36040	Ig heavy chain V-I
47	572.5	16.8	249	2 S69340	Ig heavy chain VHI
48	566	16.6	152	2 S14236	Ig gamma-1 chain C
49	401	11.7	549	2 S04845	Ig heavy chain pre
50	399.5	11.4	572	2 B46529	Ig y heavy chain (
51	389.5	11.4	627	2 S14683	Ig mu chain precu
52	374	11.0	548	2 S38864	Ig epsilon chain C
53	367.5	10.8	388	1 EHMS	Ig epsilon chain C
54	363.5	10.6	429	1 EHRT	Ig epsilon chain C
55	363.5	10.6	592	2 S25705	Ig mu chain - sh
56	360.5	10.6	448	2 S03186	Ig heavy chain C r
57	360	10.5	504	2 S00390	Ig gamma chain (cl
58	359.5	10.5	476	1 MHMSM	Ig mu chain C regi
59	359	10.5	474	2 S15590	Ig heavy chain - h
60	357.5	10.5	453	2 S37768	Ig mu chain C regi
61	357	10.5	71	2 I60082	CD4 receptor - hum
62	356	10.4	428	1 EHHU	Ig epsilon chain C
63	355.5	10.4	479	1 MHREM	Ig mu chain C regi
64	353.5	10.4	426	2 I36948	Ig epsilon chain -
65	352.5	10.3	391	1 MHHDUT	Ig mu heavy chain
66	351.5	10.3	473	1 MHHDUT	Ig mu chain C regi
67	348	10.2	455	1 MHMS	Ig mu chain C regi
68	348	10.2	455	2 A24976	Ig mu chain C regi
69	347	10.2	452	1 MHU	Ig mu chain C regi
70	346.5	10.1	423	1 EHMS	Ig epsilon chain C
71	346.5	10.1	454	1 MHY	Ig mu chain C regi
72	346	10.1	457	2 S03961	Ig mu chain C regi
73	346	10.1	458	1 MHRB	Ig heavy chain - n
74	345.5	10.1	577	2 I50731	Ig mu chain C regi
75	341.5	10.0	343	2 S25644	Ig mu chain C regi
76	340	10.0	450	1 MHG	Ig mu chain C regi
77	331.5	9.7	112	2 B30503	Ig gamma-2a chain
78	326.5	9.6	684	2 S60266	novel antigen rece
79	324.5	9.5	453	2 C31933	Ig mu chain C regi
80	320.5	9.4	433	2 S31436	Ig epsilon chain -
81	306	9.0	299	1 AHRB	Ig alpha chain C r
82	305.5	8.9	99	2 S21461	Ig alpha chain C r
83	303	8.9	338	2 S09276	Ig alpha chain C r
84	303	8.9	461	1 HVRKC	Ig mu chain C regi
85	301.5	8.8	438	1 A46507	Ig alpha chain - c
86	301.5	8.8	585	2 S09270	Ig alpha chain C r
87	298.5	8.7	342	2 I47175	Ig alpha chain C r
88	298	8.7	438	1 HVRK2	Ig mu chain C regi
89	297.5	8.7	357	2 S09269	Ig alpha chain C r
90	297	8.7	347	2 S09274	Ig alpha chain C r
91	295.5	8.7	347	2 B2360	Ig alpha chain C r
92	293.5	8.6	367	1 MHCH	Ig mu chain-2 chain
93	293	8.6	367	1 HVRK1	Ig mu chain C regi
94	289.5	8.5	393	1 S09266	Ig alpha chain C r
95	285.5	8.4	352	2 S09265	Ig alpha chain C r
96	285.5	8.4	357	2 S09272	Ig alpha chain C r
97	285	8.3	343	2 I56230	Ig alpha-2 chain -
98	284.5	8.3	340	2 S12328	Ig heavy chain C r
99	284	8.3	244	2 S05500	Ig alpha-1 chain C
100	283.5	8.3	353	1 AIHU	Ig alpha-1 chain C
101	282.5	8.3	353	1 AIHU	Ig alpha-1 chain C
102	280.5	8.2	99	2 S21462	T-cell surface gly

103 278.5 8.2 474 2 150830 Ig mu chain - Lep1  
104 278 8.1 558 2 A45804 Ig mu chain C regi  
105 277 8.1 339 2 S09264 Ig alpha chain C r  
106 277 8.1 568 2 A34891 Ig heavy chain pre  
107 275.5 8.1 88 2 A30503 Ig gamma-2b chain  
108 274 8.0 360 2 S09271 Ig alpha chain C r  
109 273.5 8.0 348 2 S09273 Ig alpha chain C r  
110 271 7.9 340 1 A2HU Ig alpha-2 chain C  
111 270 7.9 370 1 HVKCS Ig mu chain C regi  
112 270 7.9 573 2 S12838 Ig mu chain precut  
113 269.5 7.8 358 2 S09268 Ig alpha chain C r  
114 266.5 7.8 357 2 S09267 Ig alpha chain C r  
115 265 7.8 454 2 A4532 Ig mu chain C regi  
116 258.5 7.6 342 2 A45966 Ig alpha chain C r  
117 257 7.5 111 2 S43148 Ig upelion chain -  
118 251.5 7.4 344 1 AHMS Ig alpha chain C r  
119 247 7.2 1005 2 T18517 Ig heavy chain - c  
120 243.5 7.1 220 2 C22360 Ig alpha-2 chain C  
121 241.5 7.1 580 2 A46538 Ig heavy chain, se  
122 239.5 7.0 335 2 S09275 Ig alpha chain C r  
123 211 6.2 448 2 A46533 Ig heavy chain C r  
124 209 6.1 115 2 I68731 Igb chain C4 regio  
125 206 6.0 402 2 S20002 Ig heavy chain, se

ALIGNMENTS

RESULT 1  
RMHTT4  
T-cell surface glycoprotein CD4 precursor [validated] - human  
N:Alternate names: T-cell surface antigen T4/Leu 3  
C:Species: Homo sapiens (man)  
C>Date: 28-May-1986 #sequence, revision 31-Dec-1988 #text, change 20-Apr-2001  
C:Accession: A90872; A32722; A34194; A53287; I54176; I54297; A02109; A30039  
R:Maddon, P.J.; Littman, D.R.; Godfrey, M.; Maddon, D.E.; Chess, L.; Axel, R.  
Cell 42, 93-104, 1985  
A:Title: The isolation and nucleotide sequence of a cDNA encoding the T cell surface pro  
A:Reference number: A90872; MUID:85254946; PMID:2990730  
A:Accession: A90872  
A:Molecule type: mRNA  
A:Residues: 1-25, 'N', 27-458 <MAD>  
A:Experimental source: clone pT4B  
R:Littman, D.R.; Maddon, P.J.; Axel, R.  
Cell 55, 541, 1988  
A:Title: Corrected CD4 sequence.  
A:Reference number: A90907; MUID:89028665; PMID:3263213  
A:Contents: annotation; revision to residue 26  
R:Camerini, D.; Seed, B.  
Cell 60, 747-754, 1990  
A:Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the vi  
A:Reference number: A32722; MUID:90182664; PMID:2107024  
A:Accession: A32722  
A:Status: nucleic acid sequence not shown; not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 26-426 428-458 <CAM>  
R:Cart, S.A.; Hemling, M.E.; Folea-Wasserman, G.; Sweet, R.W.; Anumula, K.; Barr, J.R.;  
J. Biol. Chem. 264, 21286-21295, 1989  
A:Title: Protein and carbohydrate structural analysis of a recombinant soluble CD4 recep  
A:Reference number: A34194; MUID:90078233; PMID:2592374  
A:Contents: disulfide bonds; carbohydrate-binding sites  
A:Accession: A34194  
A:Molecule type: protein  
A:Residues: 26-394 <CAR>  
R:Lederman, S.; Demartino, J.A.; Daugherty, B.L.; Foeldvari, I.; Yellin, M.J.; Cleary, A  
Mol. Immunol. 28, 1171-1181, 1991  
A:Title: A single amino acid substitution in a common African allele of the CD4 molecule  
A:Reference number: A53287; MUID:92072595; PMID:1961196  
A:Accession: A53287  
A:Status: not compared with conceptual translation  
A:Molecule type: mRNA  
A:Residues: 250-264, 'W', 266-280 <LED>  
A:Note: sequence extracted from NCBI backbone (NCBIP:68249)

R:Edwards, M.C.; Gibbs, R.A.  
Genomics 14, 590-597, 1992  
A:Title: A human dimeric protein resulting from loss of an Alu.  
A:Reference number: I54176; MUID:93052387; PMID:1330888  
A:Accession: I54176  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-72 <RES>  
A:Cross-references: GB:U47924; GB:M6525; GB:U72506; NID:g1633547; PIDN:AA16069.1; PID  
R:Hodge, T.W.; Sasse, D.R.; McDougal, J.S.  
Hum. Immunol. 30, 99-104, 1991  
A:Title: Humans with OX74-epitope deficiency have a single nucleotide base change in th  
A:Reference number: I54297; MUID:91216786; PMID:1708753  
A:Accession: I54297  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-264, 'W', 266-458 <RE2>  
A:Cross-references: GB:M35160; NID:g179143; PIDN:AA16069.1; PID:g179144  
C:Comment: Macrophage tropic strains of HIV-1 bind to a complex of chemokine (C-C) recep  
C:Genetics:  
A:Gene: GDB:CD4  
A:Cross-references: GDB:119767; OMIM:186940  
A:Map position: 12pter-12p12  
A:Intons: 16/3  
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
C:Keywords: AIDS; duplication; glycoprotein; T-cell; transmembrane protein  
F:1-25/Domain: signal sequence #status predicted <Sig>  
F:26-458/Product: T-cell surface glycoprotein CD4 #status experimental <MAT>  
F:34-111/Domain: immunoglobulin homology <IM1>  
F:136-186/Domain: immunoglobulin homology <IM2>  
F:216-299/Domain: immunoglobulin homology <IM3>  
F:321-372/Domain: immunoglobulin homology <IM4>  
F:397-420/Domain: transmembrane #status predicted <TM>  
F:421-458/Domain: intracellular #status predicted <INT>  
F:41-109, 155-184, 328-370/disulfide bonds: #status experimental  
F:296,325/Binding site: carbohydrate (asn) (covalent) #status experimental

Query Match 59.0%; Score 2015; DB 1; Length 458;  
Best Local Similarity 99.5%; Pred. No. 3, 1e-111;  
Matches 392; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1 MNRGVFRLILVLTALPAATQGNKVLGKKGDVETCTASQKKSIOFHKNSNOIK 60  
Db 1 MNRGVFRLILVLTALPAATQGNKVLGKKGDVETCTASQKKSIOFHKNSNOIK 60  
QY 61 ILGNQGSFLTKGSPSKINDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120  
Db 61 ILGNQGSFLTKGSPSKINDRADSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEVOL 120  
QY 121 LVFGLTANSDTHLLOQOSLTLTLESPGSSPSVOCHSPGKNTQGGKTLVSQLELDQSG 180  
Db 121 LVFGLTANSDTHLLOQOSLTLTLESPGSSPSVOCHSPGKNTQGGKTLVSQLELDQSG 180  
QY 181 TWTCVTLQNOQKVEFKIDIVLAFOKASSIVYKKEGOVEFSPPLAFTVEKLTGSGELMW 240  
Db 181 TWTCVTLQNOQKVEFKIDIVLAFOKASSIVYKKEGOVEFSPPLAFTVEKLTGSGELMW 240  
QY 241 QAEARASSKSWITFDLKNKEVSKYRTQPKLOMGKPLHLTLPLPALPOYAGSSGULTLA 300  
Db 241 QAEARASSKSWITFDLKNKEVSKYRTQPKLOMGKPLHLTLPLPALPOYAGSSGULTLA 300  
QY 301 LEAKTKLHOEVNVLVVRATOLQKNTLCEVWGPTSPKMLSLKLEKKAIVSKREKPVWV 360  
Db 301 LEAKTKLHOEVNVLVVRATOLQKNTLCEVWGPTSPKMLSLKLEKKAIVSKREKPVWV 360  
QY 361 LNPBAGMOCILSDSGVLLSNIKVLPMTSTPV 394  
Db 361 LNPBAGMOCILSDSGVLLSNIKVLPMTSTPV 394

RESULT 2  
RMCTT4  
T-cell surface glycoprotein CD4 - chimpanzee

N:Alternate names: T-cell surface antigen T4/Leu 3  
 C:Species: Pan troglodytes (chimpanzee)  
 C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
 R:Cametini, D.; Seed, B.  
 C:Accession: B32722; A46534  
 A:Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the v  
 Cell 60, 747-754, 1990  
 A:Reference number: A32722; MUID:90182664; PMID:2107024  
 A:Accession: B32722  
 A:Molecule type: mRNA  
 A:Residues: 1-432 <CM>  
 A:Cross-references: GB:M31135  
 R:Fomsgard, A.; Hirsch, V.M.; Johnson, P.R.  
 Eur. J. Immunol. 22, 2973-2981, 1992  
 A:Title: Cloning and sequences of primate CD4 molecules: diversity of the cellular recep  
 A:Reference number: A46534; MUID:93049640; PMID:1425921  
 A:Accession: A46534  
 A:Status: not compared with conceptual translation  
 A:Molecule type: mRNA  
 A:Residues: 3-399 <POM>  
 A>Note: Sequence extracted from NCBI backbone (NCBIP:118332)  
 C:Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells  
 C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
 C:Keywords: duplication; glycoprotein; T-cell; transmembrane protein  
 F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>  
 F:1-371/Domains: extracellular #status predicted <EXT>  
 F:9-86/Domains: immunoglobulin homology <IM1>  
 F:111-161/Domains: immunoglobulin homology #status atypical <IM2>  
 F:191-274/Domains: immunoglobulin homology <IM3>  
 F:296-347/Domains: immunoglobulin homology <IM4>  
 F:372-395/Domains: transmembrane #status predicted <TM>  
 F:396-432/Domains: intracellular #status predicted <INT>  
 F:16-84,130-159,303-345/Disulfide bonds: #status predicted  
 F:271,300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 54.7%; Score 1867; DB 1; Length 432;  
 Best Local Similarity 98.6%; Pred. No. 1.4e-102;  
 Matches 363; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY	27	KVVLGGKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL	86
DB	2	KVVLGGKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL	61
QY	87	WDQGNFPLITKNIKIESDPTICEVEDQKEVOLLVGLTANSTHLLQGSLLTLESP	146
DB	62	WDQGNFPLITKNIKIESDPTICEVEDQKEVOLLVGLTANSTHLLQGSLLTLESP	121
QY	147	PGSSPSVQCRSPRGKNIQSGKTLVSQLELDSDGTWCTVLOKQKVEFKIDIVLAFOK	206
DB	122	PGSSPSVQCRSPRGKNIQSGKTLVSQLELDSDGTWCTVLOKQKVEFKIDIVLAFOK	181
QY	207	ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSKSMITFDLKNKEVSXRV	266
DB	182	ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSKSMITFDLKNKEVSXRV	241
QY	267	TODPKLQMGKLLPLHLLTPALPOYAGSGNLTALAEKTKLQHEVNLVYMRATQLOKNTL	326
DB	242	TODPKLQMGKLLPLHLLTPALPOYAGSGNLTALAEKTKLQHEVNLVYMRATQLOKNTL	301
QY	327	TCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLPNPAQMWOCCLSDSGOVLLESNIKV	386
DB	302	TCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLPNPAQMWOCCLSDSGOVLLESNIKV	361
QY	387	LPTWSTPV 394	
DB	362	LPTWSTPV 369	

RESULT 3  
 RMM074  
 T-cell surface glycoprotein CD4 - rheus macaque  
 N:Alternate names: T-cell surface antigen T4/Leu 3  
 C:Species: Macaca mulatta (rheus macaque)

C>Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 16-Jul-1999  
 C:Accession: C32722  
 R:Cametini, D.; Seed, B.  
 C:Accession: B32722  
 A:Title: A CD4 domain important for HIV-mediated syncytium formation lies outside the v  
 Cell 60, 747-754, 1990  
 A:Reference number: A32722; MUID:90182664; PMID:2107024  
 A:Accession: C32722  
 A:Molecule type: mRNA  
 A:Residues: 1-432 <CM>  
 A:Cross-references: GB:M31134  
 C:Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells  
 C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
 C:Keywords: duplication; glycoprotein; T-cell; transmembrane protein  
 F:1-432/Product: T-cell surface glycoprotein CD4 #status predicted <MAT>  
 F:1-371/Domains: extracellular #status predicted <EXT>  
 F:9-86/Domains: immunoglobulin homology <IM1>  
 F:111-161/Domains: immunoglobulin homology #status atypical <IM2>  
 F:180-293/Domains: immunoglobulin homology <IM3>  
 F:296-347/Domains: immunoglobulin homology <IM4>  
 F:372-395/Domains: transmembrane #status predicted <TM>  
 F:396-432/Domains: intracellular #status predicted <INT>  
 F:16-84,130-159,303-345/Disulfide bonds: #status predicted  
 F:271,300/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 50.6%; Score 1729; DB 1; Length 432;  
 Best Local Similarity 90.6%; Pred. No. 1.9e-94;  
 Matches 334; Conservative 16; Mismatches 18; Indels 0; Gaps 0;

QY	27	KVVLGGKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL	86
DB	2	KVVLGGKGDVLTCTASQKKSIOFHWKNSNQIKILGNQGSFLTKGPSKLNDRADSRSL	61
QY	87	WDQGNFPLITKNIKIESDPTICEVEDQKEVOLLVGLTANSTHLLQGSLLTLESP	146
DB	62	WDQGNFPLITKNIKIESDPTICEVEDQKEVOLLVGLTANSTHLLQGSLLTLESP	121
QY	147	PGSSPSVQCRSPRGKNIQSGKTLVSQLELDSDGTWCTVLOKQKVEFKIDIVLAFOK	206
DB	122	PGSSPSVQCRSPRGKNIQSGKTLVSQLELDSDGTWCTVLOKQKVEFKIDIVLAFOK	181
QY	207	ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSKSMITFDLKNKEVSXRV	266
DB	182	ASSIVYKKEGEQVEFSPFLAFTVEKLTGSGELMWQERASSKSMITFDLKNKEVSXRV	241
QY	267	TODPKLQMGKLLPLHLLTPALPOYAGSGNLTALAEKTKLQHEVNLVYMRATQLOKNTL	326
DB	242	TODPKLQMGKLLPLHLLTPALPOYAGSGNLTALAEKTKLQHEVNLVYMRATQLOKNTL	301
QY	327	TCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLPNPAQMWOCCLSDSGOVLLESNIKV	386
DB	302	TCEVWGPTSPKMLSLKLENKAKVSKREKPVWVLPNPAQMWOCCLSDSGOVLLESNIKV	361
QY	387	LPTWSTPV 394	
DB	362	LPTWSTPV 369	

RESULT 4  
 GHU  
 Ig gamma-1 chain C region - human  
 C:Species: Homo sapiens (man)  
 C>Date: 31-Jan-1981 #sequence\_revision 18-Aug-1982 #text\_change 16-Jul-1999  
 C:Accession: A93433; S36867; S3887; B90563; A90564; B91668; A91723; A02146  
 R:Ellison, J.W.; Berson, B.J.; Hood, L.E.  
 Nucleic Acids Res. 10, 4071-4079, 1982  
 A:Title: The nucleotide sequence of a human immunoglobulin C-gamma1 gene.  
 A:Reference number: A93433; MUID:82274238; PMID:6287432  
 A:Accession: A93433  
 A:Molecule type: DNA  
 A:Residues: 1-330 <EHL>  
 A:Cross-references: EMBL:217370  
 A>Note: this sequence has the Gln(17) allelic marker, 97-Lys, and the Gln(1) markers,  
 A:Note: Lys-330 is removed after translation

R,Harris, L.J.  
 submitted to the EMBL Data Library, October 1992  
 A:Reference number: S33904  
 A:Accession: S36861  
 A:Molecule type: DNA  
 A:Residues: 2-330 <HAR>  
 A:Cross-references: EMBL:Z17370  
 R:Yokoyama, N.; Ueda, S.; Obata, M.; Nikaide, T.; Nakai, S.; Honjo, T.  
 Cell 29, 671-679, 1982  
 A:Title: Structure of human immunoglobulin gamma genes: implications for evolution of a  
 A:Reference number: S33887; MUID:83001943; PMID:6811139  
 A:Accession: S33887  
 A:Molecule type: DNA  
 A:Residues: 88-113;235-330 <TAK>  
 A:Cross-references: EMBL:Z17370  
 R:Cunningham, B.A.; Rutishauser, U.; Gall, W.E.; Gottlieb, P.D.; Maxdall, M.J.; Edelman,  
 Biochemistry 9, 3161-3170, 1970  
 A:Title: The covalent structure of a human gammaG-immunoglobulin. VII. Amino acid sequen  
 A:Reference number: A90563; MUID:71064024; PMID:5489771  
 A:Accession: myeloma protein Eu  
 A:Accession: B90563  
 A:Molecule type: protein  
 A:Residues: 1-96, 'R', 98-135 <CUN>  
 A:Note: this sequence has the G1m(3) marker, 97-Arg  
 R:Rutishauser, U.; Cunningham, B.A.; Bennett, C.; Konigsberg, W.H.; Edelman, G.M.  
 Biochemistry 9, 3171-3181, 1970  
 A:Title: The covalent structure of a human gammaG-immunoglobulin. VIII. Amino acid sequen  
 A:Reference number: A90564; MUID:71064025; PMID:5530842  
 A:Accession: Eu  
 A:Accession: A90564  
 A:Molecule type: protein  
 A:Residues: 136-154, 'Q', 156-165, 'Q', 167-176, 'Q', 178-194, 'N', 196-197, 'D', 199-238, 'E', 240,  
 A:Note: this sequence has the G1m(non-1) markers, 239-Glu and 241-Met  
 R:Ponsing, H.; Hilschmann, N.  
 Hoppe-Seyler's Z. Physiol. Chem. 357, 1571-1604, 1976  
 A:Title: Die Primarstruktur eines monoklonalen IgG1-Immunglobulins (Myelomprotein Nie),  
 Igen Primerstruktur.  
 A:Reference number: A91668; MUID:77070269; PMID:826475  
 A:Accession: myeloma protein Nie  
 A:Accession: B91668  
 A:Molecule type: protein  
 A:Residues: 1-34, 'Q', 36-96, 'K', 98-115, 'Q', 117-197, 'D', 199-238, 'D', 240, 'L', 242-268, 'E', 27  
 A:Note: this sequence has the G1m(17) and G1m(1) markers  
 R:Schmidt, W.E.; Jung, H.D.; Palm, W.; Hilschmann, N.  
 Hoppe-Seyler's Z. Physiol. Chem. 364, 713-747, 1983  
 A:Title: Die Primarstruktur des kristallisierten monoklonalen Immunglobulins IgG1 KOI  
 A:Reference number: A91723; MUID:83289131; PMID:6884994  
 A:Accession: myeloma protein KOI; disulfide bonds  
 A:Accession: A91723  
 A:Molecule type: protein  
 A:Residues: 1-96, 'R', 98-197, 'D', 199-238, 'E', 240, 'W', 242-266, 'D', 268-271, 'D', 273-330 <SCH  
 A:Note: this sequence has the G1m(3) and G1m(non-1) markers  
 R:Gall, W.E.; Edelman, G.M.  
 Biochemistry 9, 3188-3196, 1970  
 A:Title: The covalent structure of a human gammaG-immunoglobulin. X. Intrachain disulfid  
 A:Reference number: A90565; MUID:71064027; PMID:4923144  
 A:Accession: disulfide bonds  
 A:Contents: annotation; disulfide bonds  
 R:Drexler, L.; Schwarz, J.; Reichel, W.; Hilschmann, N.  
 Hoppe-Seyler's Z. Physiol. Chem. 357, 1515-1540, 1976  
 A:Title: Rule of antibody structure. The primary structure of monoclonal IgG1 immunoglob  
 embionide cleavage products, and the disulfide bridges.  
 A:Reference number: A91667; MUID:77070267; PMID:1002129  
 A:Contents: annotation; disulfide bonds  
 C:Genetics:  
 A:Gene: CDB:IGHG1  
 A:Cross-references: GDB:120085; OMIM:147100  
 A:Map position: 14q32.33-14q32.33  
 A:Features: 99/1, 114/1, 224/1  
 C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap  
 hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into la  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin  
 F:20-85/Domain: immunoglobulin homology <IM2>

F:137-206/Domain: immunoglobulin homology <IM2>  
 F:243-310/Domain: immunoglobulin homology <IM2>  
 F:27-83,144-204,250-308/Disulfide bonds: #status experimental  
 F:103/Disulfide bonds: interchain (to light chain) #status experimental  
 F:109,112/Disulfide bonds: interchain (to heavy chain) #status experimental  
 F:180/Binding site: carbohydrate (Aan) (covalent) #status experimental  
 Query Match 35.5%; Score 121.5; DB 1; Length 330;  
 Best Local Similarity 97.8%; Pred. No. 3.6e-64;  
 Matches 224; Conservative 0; Mismatches 0; Indels 5; Gaps 1;  
 400 EPKSCDKHTC-----PELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 454  
 99 EPKSCDKHTCPCPCAPPELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 158  
 455 NMVVDGVEVHNAKTRERQYNSTYRVSVLTVLDHQMVGKVKCKVSNKALPAPIEKT 514  
 159 NMVVDGVEVHNAKTRERQYNSTYRVSVLTVLDHQMVGKVKCKVSNKALPAPIEKT 218  
 515 ISKAKQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTP 574  
 219 ISKAKQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTP 278  
 575 PVLDSGSEFLYSKLTVDKSRMVGQGVVFSCVMEALHNHYTQKSLSLSPG 625  
 279 PVLDSGSEFLYSKLTVDKSRMVGQGVVFSCVMEALHNHYTQKSLSLSPG 329  
 RESULT 5  
 S69339  
 Ig heavy chain V region precursor - human  
 C:Species: Homo sapiens (man)  
 C:Date: 19-Mar-1997 #sequence revision 19-Mar-1997 #ext\_change 01-Dec-2000  
 C:Accession: S69339; S72664  
 R:Khamlichi, A.A.; Aucouturier, P.; Preud'homme, J.L.; Cogne, M.  
 Eur. J. Biochem. 229, 54-60, 1995  
 A:Title: Structure of abnormal heavy chains in human heavy-chain-deposition disease.  
 A:Reference number: S69339; MUID:95262687; PMID:7744049  
 A:Accession: S69339  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-374 <KHA>  
 A:Cross-references: EMBL:X81695  
 R:Khamlichi, A.A.  
 submitted to the EMBL Data Library, September 1994  
 A:Reference number: S72664  
 A:Accession: S72664  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-140, 'C', 142-374 <KH2>  
 A:Cross-references: EMBL:X81695  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 Query Match 35.3%; Score 1205.5; DB 2; Length 374;  
 Best Local Similarity 97.0%; Pred. No. 9.6e-64;  
 Matches 224; Conservative 2; Mismatches 0; Indels 5; Gaps 1;  
 400 EPKSCDKHTC-----PELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 454  
 143 EPKSCDKHTCPCPCAPPELLGSPVFLFPKPKDTLMISRTPEVTCVVVDVSHEDPEVKF 202  
 455 NMVVDGVEVHNAKTRERQYNSTYRVSVLTVLDHQMVGKVKCKVSNKALPAPIEKT 514  
 203 NMVVDGVEVHNAKTRERQYNSTYRVSVLTVLDHQMVGKVKCKVSNKALPAPIEKT 262  
 515 ISKAKQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTP 574  
 263 ISKAKQPREPQVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVEWESNGQPENNYKTP 322  
 575 PVLDSGSEFLYSKLTVDKSRMVGQGVVFSCVMEALHNHYTQKSLSLSPG 625  
 323 PVLDSGSEFLYSKLTVDKSRMVGQGVVFSCVMEALHNHYTQKSLSLSPG 373

## RESULT 6

S31866

Ig gamma-1 chain C region - synthetic

C/Species: synthetic

A/Note: Homo sapiens (man) gene engineered and expressed in Escherichia coli

C/Date: 06-Jan-1995 #sequence\_revision 17-Mar-1997 #text\_change 19-May-2000

C/Accession: S31866

R/Fltpula, D.

Submitted to the EMBL Data Library, February 1993

A/Description: Screening method for protein-protein interactions of cloned gene products.

A/Reference number: S31866

A/Accession: S31866

A/Molecule type: mRNA

A/Residues: 1-255 &lt;FLP&gt;

A/Cross-references: EMBL:X70421; NID:g33068; PIDN:CAA4866.1; PID:g33068

C/Keywords: Immunoglobulin

F/1-22/Region: Escherichia coli outer membrane protein A precursor

F/23-255/Region: human Ig gamma-1 chain C region

Query Match 35.3%; Score 1203.5; DB 4; Length 255;

Best Local Similarity 97.4%; Pred. No. 7.7e-64;

Matches 225; Conservative 0; Mismatches 1; Indels 5; Gaps 1;

Oy 400 EPKSCDKTHTC-----PELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKF 454  
|||||  
Db 24 ESKSCDKTHTCPKCAPPELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKF 83  
|||||  
Oy 455 NMVYDGEVHNAAKTPREEOYNSTRVVSVLTVLHODMLNGEKYCKVSNKALPAPIEKT 514  
|||||  
Db 84 NMVYDGEVHNAAKTPREEOYNSTRVVSVLTVLHODMLNGEKYCKVSNKALPAPIEKT 143  
|||||  
Oy 515 ISKAKGOREPOVYTLTPSRDELTKNOVSLTCLVKGFPSPDIAVWESNGOPENNYKTTP 574  
|||||  
Db 144 ISKAKGOREPOVYTLTPSRDELTKNOVSLTCLVKGFPSPDIAVWESNGOPENNYKTTP 203  
|||||  
Oy 575 PVLDSGSPFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKSLSPG 625  
|||||  
Db 204 PVLDSGSPFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKSLSPG 254  
|||||

## RESULT 7

PT0207

Ig gamma chain C region - chimpanzee

C/Species: Pan troglodytes (chimpanzee)

C/Date: 23-Nov-1991 #sequence\_revision 23-Nov-1991 #text\_change 16-Jul-1999

C/Accession: PT0207

R/Enrich, P.H.; Moustafa, Z.A.; Oestberg, L.

Mol. Immunol. 28, 319-322, 1991

A/Title: Nucleotide sequence of chimpanzee FC and hinge regions.

A/Reference number: PT0207; MUID:91287716; PMID:2062315

A/Accession: PT0207

A/Molecule type: mRNA

A/Residues: 1-234 &lt;EHR&gt;

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: immunoglobulin

F/48-117/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 34.3%; Score 1171.5; DB 2; Length 234;  
Best Local Similarity 96.9%; Pred. No. 5.3e-62;  
Matches 218; Conservative 1; Mismatches 1; Indels 5; Gaps 1;

Oy 400 EPKSCDKTHTC-----PELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKF 454  
|||||  
Db 10 EPKSCDTHTCPKCAPPELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDVSHEDPEVKF 69  
|||||  
Oy 455 NMVYDGEVHNAAKTPREEOYNSTRVVSVLTVLHODMLNGEKYCKVSNKALPAPIEKT 514  
|||||  
Db 70 NMVYDGEVHNAAKTPREEOYNSTRVVSVLTVLHODMLNGEKYCKVSNKALPAPIEKT 129  
|||||  
Oy 515 ISKAKGOREPOVYTLTPSRDELTKNOVSLTCLVKGFPSPDIAVWESNGOPENNYKTTP 574  
|||||  
Db 130 ISKAKGOREPOVYTLTPSRDELTKNOVSLTCLVKGFPSPDIAVWESNGOPENNYKTTP 189  
|||||

Oy 575 PVLDSGSPFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKS 619  
|||||  
Db 190 PVLDSGSPFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKS 234  
|||||

## RESULT 8

A23511

Ig gamma-3 chain C region (allotype G3m(b)) - human

C/Species: Homo sapiens (man)

C/Date: 28-Dec-1987 #sequence\_revision 28-Dec-1987 #text\_change 23-Jul-1999

C/Accession: A23511

R/Huck, S.; Fort, P.; Crawford, D.H.; Lefranc, M.P.; Lefranc, G.

Nucleic Acids Res. 14, 1779-1789, 1986

A/Title: Sequence of a human immunoglobulin gamma 3 heavy chain constant region gene: c

A/Reference number: A23511; MUID:86148507; PMID:3081877

A/Accession: A23511

A/Molecule type: DNA

A/Residues: 1-377 &lt;HUC&gt;

A/Cross-references: GB:X03604; GB:M12958; NID:g33070; PIDN:CAA27268.1; PID:g577056

C/Genetics:

A/Gene: IGHG3

A/Cross-references: GDB:119339; OMIM:147120

A/Map position: 14q32.33-14q32.33

A/Intons: 98/3; 115/3; 130/3; 145/3; 160/3; 270/3

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: immunoglobulin

F/20-85/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 34.1%; Score 1163; DB 2; Length 377;  
Best Local Similarity 90.0%; Pred. No. 3.1e-61;  
Matches 216; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

Oy 392 TPVPCP-APPKSCDKTHTC-----PELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDV 445  
|||||  
Db 137 TPVPCP-APPKSCDKTHTC-----PELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDV 196  
|||||  
Oy 446 SHEDPEVKFMVYDGEVHNAAKTPREEOYNSTRVVSVLTVLHODMLNGEKYCKVSNK 505  
|||||  
Db 197 SHEDPEVKFMVYDGEVHNAAKTPREEOYNSTRVVSVLTVLHODMLNGEKYCKVSNK 256  
|||||  
Oy 506 ALPAPIEKTISKAKGOREPOVYTLTPSRDELTKNOVSLTCLVKGFPSPDIAVWESNGO 565  
|||||  
Db 257 ALPAPIEKTISKAKGOREPOVYTLTPSRDELTKNOVSLTCLVKGFPSPDIAVWESNGO 316  
|||||  
Oy 566 PENNYKTTPVLDSDGSPFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKSLSPG 625  
|||||  
Db 317 PENNYKTTPVLDSDGSPFLYSKLTVDKSRMOQGNVSCVMHEALHNHTOKSLSPG 376  
|||||

## RESULT 9

A60764

Ig gamma-3 chain C region, form LAT - human

C/Species: Homo sapiens (man)

C/Date: 14-May-1993 #sequence\_revision 14-May-1993 #text\_change 16-Jul-1999

C/Accession: A60764

R/Huck, S.; Lefranc, G.; Lefranc, M.P.

Immunogenetics 30, 250-257, 1989

A/Title: A human immunoglobulin IGHG3 allele (Gmb0, b1, c3, c5, u) with an IGHG4 conver

A/Reference number: A60764; MUID:90007613; PMID:2571587

A/Accession: A60764

A/Status: preliminary

A/Molecule type: DNA

A/Residues: 1-377 &lt;HUC&gt;

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: immunoglobulin

F/20-85/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 34.0%; Score 1161; DB 2; Length 377;  
Best Local Similarity 90.0%; Pred. No. 4e-61;  
Matches 216; Conservative 8; Mismatches 10; Indels 6; Gaps 2;

Oy 392 TPVPCP-APPKSCDKTHTC-----PELLGSPSVFLPPPKKDTLMISRTPEVTCVVVDV 445  
|||||

```

Db      137 TTPPCRCRCEPKSCDTPPCRCRCPABELLGSGSVFLFPKPKDITLMISTRTPEVTCVVDV 196
QY      446 SHEDPEVKFNWYDGVGVHNAKTKPREEOYNSTYRVSVLTVLHQDLNKEVEKCKVSNK 505
Db      197 SHEDPEVQFKWYDGVGVHNAKTKPREEOYNSTFRVSVLTVLHQDLNKEVEKCKVSNK 256
QY      506 ALPAPLEKTIKSKAGQPREPOVYTLTPSRDELTKNOVSLTCLVKGFYPSDIAVEMESNQ 565
Db      257 ALPAPLEKTIKSKAGQPREPOVYTLTPSRDEEMTKNOVSLTCLVKGFYPSDIAVEMESNQ 316
QY      566 PENNYKTPPVLDSDSFFLYSKLTVDKSRMQQGNFSCSVMEALAHNYTQKSLSLSPG 625
Db      317 PENNYKTPPVLDSDSFFLYSKLTVDKSRMQQGNFSCSVMEALAHNYTQKSLSLSPG 376

```

## RESULT 10

```

A:Accession: A46254
C:Species: Oryctolagus cuniculus (domestic rabbit)
C:Date: 21-Sep-1993 #sequence, revision 18-Nov-1994 #text_change 21-Jan-2000
C:Accession: A46254
R:Hague, B.F.; Sawadikoso, S.; Brown, T.J.; Lee, K.; Recker, D.P.; Kindt, T.J.
Proc. Natl. Acad. Sci. U.S.A. 89, 7963-7967, 1992
A>Title: CD4 and its role in infection of rabbit cell lines by human immunodeficiency vi
A:Reference number: A46254; PMID:92390370; PMID:1518821
A:Accession: A46254
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-459 <HAG>
A:Cross-references: GB:M92840; NID:9164871; PIN:AAA31198.1; PID:9164872
A>Note: sequence extracted from NCBI backbone (NCBIN:112731; NCBI:P.112733)
C:Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology
F:322-372/Domain: immunoglobulin homology <IMM>

```

```

Query Match      33.7%; Score 1150.5; DB 2; Length 459;
Best Local Similarity 57.2%; Pred. No. 2,1e-60;
Matches 241; Conservative 65; Mismatches 90; Indels 25; Gaps 6;

```

```

QY      1 MNRGVPRHLVLVQLALPAATQGNKVLGKKGDTVELTCTASQKKSIOFHMKNSNQK 60
Db      1 MNRRIYFOCLLVLPALPLPAATMGKTVVGKAGATVELPCQSSQKNSVFNKMHANQVK 60
QY      61 ILGNQ-----SFLTKGSRKLNDRADSRSLMDQNPFLIKLKIENSDPTICEVDOKE 116
Db      61 ILGNQSSSSSSFWLKGNSPLSNVESKKNWMDGSPFLVKDLRMDSGTYYICEVGDCKM 120
QY      117 EVOLLVFGLTANSRDLHLOQSITLTLSPSSPSVQCRPRGKNIQGGKTLVSQLEL 176
Db      121 EVELVLPRLTANRNLRLHQSITLTLSEGSVGSPSVQWSPENKIIEGPTCSMKRL 180
QY      177 QDSGTWTCTV-LQNKQKVEFKIDIVLAFQKASSIYKKEGDEVERSPFLAFTVEKLTGS 235
Db      181 QDSGTWCHLSFQDQNKLELDIKIIVLGFPKASATVKKKEGDEVERSPFLNFEDESL--S 238
QY      236 GELMWQAEKASSSKSWTTPPLKNKEVSVKRVTDQPKQMKKLPINHLTPQALPOVAGSG 295
Db      239 GELMWQVAGASSSQSWVSFLEDRKVSQKILPDLKIQMSKGLPLSTTLTQALHRYAGSG 298
QY      296 NLTLALAKTKGLHGVNLYVMRATOLQKVLTCVEMGPTSPKMLSLKLENKEAKYSKRE 355
Db      299 NLTLTLD--KGLKHQGVSLVWLKTVQKNTLCEVGLIDIPKTLSTKLEDOEAKYS-ITQ 355
QY      356 KPVAVNLPEAGMOCCLISDGGVLESNIIVLPTWSTVPVCPAPEPKSCDKTHTCPPELLG 415
Db      356 KMTQVLDPKAGTWCCLISDGVLLSEKADVLAT-----GLSHQOQPTLLA 400
QY      416 G 416
Db      401 G 401

```

RESULT 11

## G3HUV1

```

Ig gamma-3 heavy chain disease proteins - human
C:Species: Homo sapiens (man)
C:Date: 31-Dec-1979 #sequence, revision 23-Oct-1991 #text_change 16-Jul-1999
C:Accession: A90442; A92219; A90198; A93915; A02149
R:Frangione, B.; Rosenwasser, E.; Prelli, F.; Franklin, E.C.
Biochemistry 19, 4304-4308, 1980
A>Title: Primary structure of human gamma3 immunoglobulin deletion mutant: gamma3 heavy-
A:Reference number: A90442; PMID:81021548; PMID:6774747
A:Accession: A90442
A:Molecule type: protein
A:Residues: 1-289 <FRA>
A>Note: the molecule is a dimer linked by 12 disulfide bonds; it has an extra interchari
A>Note: this protein lacks most of the V region and all of the CH1 region. Residue 12 c
A>Note: the sequence of residues 42-76 was taken from the reference that follows
R:Michaelson, T.E.; Frangione, B.; Franklin, E.C.
J. Biol. Chem. 252, 883-889, 1977
A>Title: Primary structure of the 'hinge' region of human IgG3. Probable quadruplication
A:Reference number: A92219; PMID:77118561; PMID:402363
A:Accession: A92219
A:Contents: normal gamma-3 chains, sequence corresponding to residues 12-97 of protein 1
A:Molecule type: protein
A:Residues: 12-97 <MIC>
A:Note: the hinge region in gamma-3 chains is about four times as long as in other gamma
A:Note: cysteines at positions 24, 27, 33, 39, 42, 48, 54, 57, 63, 69, and 72 form inter
R:Wollentein-Todell, C.; Frangione, B.; Prelli, F.; Franklin, E.C.
Biochem. Biophys. Res. Commun. 71, 907-914, 1976
A>Title: The amino acid sequence of "heavy chain disease" protein ZUC. Structure of the
A:Reference number: A90198; PMID:77021516; PMID:423945
A:Accession: A90198
A:Contents: heavy chain disease protein ZUC, partial sequence corresponding to residues
A:Molecule type: protein
A:Residues: 59-125, 'EB', 128-226, 228-289 <MOL>
A:Note: this protein lacks most of the V region, all of the CH1 region, and part of the
R:Alexander, A.; Steinmetz, M.; Barilett, D.; Frangione, B.; Franklin, E.C.; Hood, L.
Proc. Natl. Acad. Sci. U.S.A. 79, 3260-3264, 1982
A>Title: gamma heavy chain disease in man: cDNA sequence supports partial gene deletion
A:Reference number: A93915; PMID:82247835; PMID:6808505
A:Accession: A93915
A:Contents: heavy chain disease protein Omni
A:Molecule type: mRNA
A:Residues: 12-70, 72-114, 116-125, 'E', 127-133, 'L', 135-136, 'E', 138, 'Y', 140-154, 'D', 156-15
A:Note: a carboxyl-terminal lys is removed posttranslationally
A:Note: this sequence may represent an allelic form or another gamma chain subclass
C:Comment: The heavy chain disease protein wis is shown.
C:Genetics:
A:Gene: GDB:IGHG3
A:Cross-references: GDB:119339; OMIM:147120
A:Map position: 14q32.33-14q32.33
C:Superfamily: immunoglobulin C region; immunoglobulin homology
F:203-270/Domain: immunoglobulin homology <IMM>
F:1/Modified site: pyrrolidone carboxylic acid (Gln) #status experimental
F:6,140/Binding site: carbohydrate (Asn) (covalent) #status experimental

```

```

Query Match      33.5%; Score 1143; DB 1; Length 289;
Best Local Similarity 87.9%; Pred. No. 3,3e-60;
Matches 211; Conservative 13; Mismatches 10; Indels 6; Gaps 2;

```

```

QY      392 TTPPCRCRCEPKSCDTPPCRCRCPABELLGSGSVFLFPKPKDITLMISTRTPEVTCVVDV 445
Db      50 TTPPCRCRCEPKSCDTPPCRCRCPABELLGSGSVFLFPKPKDITLMISTRTPEVTCVVDV 109
QY      446 SHEDPEVKFNWYDGVGVHNAKTKPREEOYNSTYRVSVLTVLHQDLNKEVEKCKVSNK 505
Db      110 SHEDPEVQFKWYDGVGVHNAKTKPREEOYFNSTFRVSVLTVLHQDLNKEVEKCKVSNK 169
QY      506 ALPAPLEKTIKSKAGQPREPOVYTLTPSRDELTKNOVSLTCLVKGFYPSDIAVEMESNQ 565
Db      170 ALPAPLEKTIKSKAGQPREPOVYTLTPSRDEEMTKNOVSLTCLVKGFYPSDIAVEMESNQ 229

```



QY 566 PENNYKTPVLDSDGSEFLYSKLTVDKSRWQGNVSCSVMEALHNHYTKSLSPG 625  
 |||||  
 Db 230 PENNYKTPVLDSDGSEFLYSKLTVDKSRWQGNVSCSVMEALHNHYTKSLSPG 289

## RESULT 12

G2HU

IG gamma-2 chain C region - human

C:Species: Homo sapiens (man)

C:Date: 30-Apr-1991 #sequence\_revision 13-Jun-1993 #text\_change 21-Jul-2000

C:Accession: A93906; A92809; A90752; A93132; A02148

R:Ellison, J.; Hood, L.

Proc. Natl. Acad. Sci. U.S.A. 79, 1984-1988, 1982

A:Title: Linkage and sequence homology of two human immunoglobulin gamma heavy chain c

A:Reference number: A93906; MUID:82197621; PMID:6804948

A:Accession: A93906

A:Molecule type: DNA

A:Residues: 1-326 &lt;ELL&gt;

A:Cross-references: GB:V00554; GB:V00230; NID:932759; PIDN:CAB58438.1; PID:G606056

A:Note: Lys-326 is probably removed posttranslationally

R:Wang, A.C.; Tung, E.; Fudenberg, H.H.

J. Immunol. 125, 1048-1054, 1980

A:Title: The primary structure of a human IgG2 heavy chain: genetic, evolutionary, and f

A:Reference number: A92809; MUID:81007873; PMID:6774012

A:Accession: A92809

A:Molecule type: protein

A:Residues: 1-19, 'Q', 21-57, 'Z', 59, 'A', 61-193, 'D', 195-325 &lt;MAN&gt;

A:Note: Trp-156 is at or near the complement-binding site

R:Connell, G.E.; Parr, D.M.; Hofmann, T.

Can. J. Biochem. 57, 758-767, 1979

A:Title: The amino acid sequences of the three heavy chain constant region domains of a

A:Reference number: A90752; MUID:80001357; PMID:113060

A:Accession: A90752

A:Molecule type: protein

A:Residues: 1-24, 'E', 26-57, 'EV', 60-85, 132-171, 'ZZZ', 175, 'B', 177-193, 'D', 195-196, 'Q', 198-

A:Note: this sequence has since been revised

R:Hofmann, T.; Parr, D.M.

Mol. Immunol. 16, 923-925, 1979

A:Title: A note on the amino acid sequence of residues 381-391 of human immunoglobulin g

A:Reference number: A93132; MUID:80114419; PMID:118920

A:Accession: A93132

A:Molecule type: protein

A:Residues: 238-275 &lt;HOF&gt;

R:Hofmann, T.; Parr, D.M.

submitted to the Atlas, March 1980

A:Reference number: A94591

A:Contents: annotation; Zie, revisions to residues 25, 59, 60, and 264-268

A:Note: the revised sequence differs from that shown in having 60-Ala and in the amidati

ned

R:Milstein, C.; Frangione, B.

Biochem. J. 121, 217-225, 1971

A:Title: Disulphide bridges of the heavy chain of human immunoglobulin G2.

A:Reference number: A90253; MUID:7203500; PMID:4940472

A:Contents: annotation; myeloma protein Sa, disulfide bonds

R:Frangione, B.; Milstein, C.; Pink, J.R.L.

Nature 221, 145-148, 1969

A:Title: Structural studies of immunoglobulin G.

A:Reference number: A93157; MUID:68064124; PMID:5782707

A:Contents: annotation; Sa, disulfide bonds

C:Genetics:

A:Gene: GDB:IGHG2

A:Cross-references: GDB:119338; OMIM:147110

A:Map position: 14q32.33-14q32.33

C:Complex: An immunoglobulin heterotetramer subunit consists of two identical light (ka

hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 12

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: duplication; glycoprotein; heterotetramer; immunoglobulin

F:20-85/Domain: immunoglobulin homology &lt;IM1&gt;

F:133-202/Domain: immunoglobulin homology &lt;IM2&gt;

F:239-306/Domain: immunoglobulin homology &lt;IM3&gt;

F:14/Disulfide bonds: interchain (to light chain) #status experimental  
 F:27-83,140-200,246-304/Disulfide bonds: #status experimental  
 F:102,103,106,109/Disulfide bonds: interchain (to heavy chain) #status experimental  
 F:176/Binding site: carbohydrate (Asn) (covalent) #status predicted

## Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

Matches 206; Conservative 8; Mismatches 4; Indels 13; Gaps 1;

Query Match

32.7%; Score 115.5; DB 1; Length 326;

Best Local Similarity 89.2%; Pred. No. 1.6e-50;

QY 455 NMVVDGEVHNATKPREQOYNSTYRVSVLTVLHODMLNGKEYCKCKNSKALPAPIEKT 514  
 |||||  
 Db 156 NMVVDGEVHNATKPREQOYNSTYRVSVLTVLHODMLNGKEYCKCKNSKALPAPIEKT 215  
 |||||  
 QY 515 ISKAKQPREPOVYTLPPSRDELTKNQVSLTCLVKGFPSDIAVEMESNQGPENNYKTP 574  
 |||||  
 Db 216 ISKAKQPREPOVYTLPPSGEEMTKNQVSLTCLVKGFPSDIAVEMESNQGPENNYKTP 275  
 |||||  
 QY 575 PVLDSGSEFLYKSLTYDKSRMOQGVNFGSCVNHKALHNYTQKSLSLSPG 625  
 |||||  
 Db 276 PVLDSGSEFLYKSLTYDKSRMOQGVNFGSCVNHKALHNYTQKSLSLSPG 326  
 |||||

## RESULT 14

S30193  
 T-cell surface glycoprotein CD4 - dog  
 C/Species: Canis lupus familiaris (dog)  
 C/Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 21-Jul-2000  
 C/Accession: S30193  
 R/Mide, K.F.; Conner, G.E.; Mintz, D.H.; Alejandro, R.  
 Biochim. Biophys. Acta 1172, 315-318, 1993  
 A/Title: Primary structure of the canine CD4 antigen.  
 A/Reference number: S30193; MUID:93192324; PMID:7916632  
 A/Accession: S30193  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-432 <MIL>  
 A/Cross-references: EMBL:X68565; NID:9288652; PIDN:CA837664.1; PID:94467377  
 C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
 C/Keywords: glycoprotein  
 F/202-311/Domain: immunoglobulin homology <IM>

Query Match 32.2%; Score 1099; DB 2; Length 432;  
 Best Local Similarity 57.7%; Pred. No. 2, 1e-57;  
 Matches 226; Conservative 62; Mismatches 86; Indels 18; Gaps 6;

QY 12 LVVLQALLPAATQGNKVVLGKGGPTVELTCTASOKKSIOFHWKNSNOIKILNGQSFRTK 71  
 |||||  
 Db 1 LMLQVWLPVAVTPEYREVVLGKAGDAVELPCQTSOKKIHFNMRDSSWVQTLNGQSFRTW 60  
 |||||  
 QY 72 GPEKLRADRSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEVQLVFGILTA---- 127  
 |||||  
 Db 61 GSRRLKRVESKKNLMQGSFPLVTKDLVADSGIYECDF-DKQVEILLVFLTKMDS 119  
 |||||  
 QY 128 ----NSDTHLLQGOSITLTLEBPPGSSPSVQCSPRGKNIQGGKTLVSQLELDQSGTW 182  
 |||||  
 Db 120 GSSSGSSNIRLLOGQQLTLTLENPSGSSPSVQWKGPKNKGQONLSLWPELODQGTW 179  
 |||||  
 QY 183 TCTVLQNOQKVEFKIDIVLAFOKASSIVYKKEGOVEFSPPLAFVTEKLTGSGELMWQA 242  
 |||||  
 Db 180 TCTIISQSQKTEVERINIVLAFOKAVNTFYAREDOVEFSPPLAFVTEKLTGSGELMWQA 237  
 |||||  
 QY 243 ERASSSKSWITFDLKNKEVSKRVTDPKLQMGKKLPLHLTLPOALPOVAGSNTLTAL 302  
 |||||  
 Db 238 QGASSSLTWISFTLENNKLSMKEMHAPLKLQMKESPLRLRTLQVLSRVAGSGLTLNL- 296  
 |||||  
 QY 303 AKTGKLGHOEVNLYVMRATQLOKNTLTCVWGPSTPKMLSLKENKEAKYSKREKPVWLN 362  
 |||||  
 Db 297 AK-GTYQEVNLYVMRANSSQNNMLTCEVLPTSELTSLINKLKOAAKYSKQKLLWVVD 355  
 |||||  
 QY 363 PEAGMOCILSDSGOVLLESNIKVLPTWSPNV 394  
 |||||  
 Db 356 PEAGTMOCLLSDKDKVLAASSLVN---SSPV 383  
 |||||

## RESULT 15

A27449  
 T-cell surface glycoprotein CD4 precursor - rat  
 N/Alternate names: W3/25 antigen  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 21-May-1988 #sequence\_revision 21-May-1988 #text\_change 21-Jan-2000  
 C/Accession: A27449; A35433  
 R/Clark, S.J.; Jefferies, W.A.; Barclay, A.N.; Gagnon, J.; Williams, A.F.

Proc. Natl. Acad. Sci. U.S.A. 84, 1649-1653, 1987  
 A/Title: Peptide and nucleotide sequences of rat CD4 (W3/25) antigen: evidence for deriv  
 A/Reference number: A27449; MUID:87175555; PMID:3104900  
 A/Accession: A27449  
 A/Molecule type: mRNA  
 A/Residues: 1-457 <CLAA>  
 A/Cross-references: GB:M15768; NID:9203387; PIDN:AAA40901.1; PID:9203388  
 R/Davis, S.J.; Ward, H.A.; Pukavec, M.J.; Willie, A.C.; Williams, A.F.; Barclay, A.N.  
 U. Biol. Chem. 265, 10410-10418, 1990  
 A/Title: High level expression in Chinese hamster ovary cells of soluble forms of CD4 T  
 A/Reference number: A35433; MUID:90285164; PMID:2113054  
 A/Contents: annotation  
 C/Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
 C/Keywords: glycoprotein; membrane protein; surface antigen  
 F/219-300/Domain: immunoglobulin homology <IM>

Query Match 29.3%; Score 1000.5; DB 2; Length 457;  
 Best Local Similarity 50.1%; Pred. No. 1, 4e-51;  
 Matches 213; Conservative 64; Mismatches 127; Indels 21; Gaps 6;

QY 1 NMRGVPFRLL-LVVLQALLPAATQGNKVVLGKGGPTVELTCTASOKKSIOFHWKNSNO 58  
 |||||  
 Db 1 MCRGSEFRLTLVLLQSLVLTQGTVLGEGSAELPCESTSRSSASAPWMSDQ 60  
 |||||  
 QY 59 IKILNGQSEFLTKGPEKLRADRSRSLMDQGNFPLIKNLKIEDSDTYICEVEDQKEEV 118  
 |||||  
 Db 61 KTIILGKKNLTKGSELTSRPSDRKNAMERGSFPLIKNLKIEDSDTYICEVEDQKEEV 120  
 |||||  
 QY 119 QLLVFGLTANSDTHLLQGSITLTLES-PPGSSPSVQCSPRGKNIQGGKTLVSQLELDQ 177  
 |||||  
 Db 121 ELWVFVTEVTPGTRILLQGSITLTLDNSNPKVSDPTECHGKSNIVKDSKAFSTSLRIQ 180  
 |||||  
 QY 178 DSGTWTCTVLQNOQKVEFKIDIVLAFOKASSIVYKKEGOVEFSPPLAFVTEKLTGSGE 237  
 |||||  
 Db 181 DSGWVCTVTLQKQKSFPMKLSVLFQASTSTAYVSEBSAEFSPPLVGEESTL--QGE 238  
 |||||  
 QY 238 LWMQERASSSKSWITFDLKNKEVSKRVTDPKLQMGKKLPLHLTLPOALPOVAGSNTL 297  
 |||||  
 Db 239 LMKAKASQSSWITFSLKNQKVSQKSTSNPKQLSTPLTLOIPQVSLQFAGSNTL 298  
 |||||  
 QY 298 TLAEAKTGKLGHOEVNLYVMRATQLOKNTLTCVWGPSTPKMLSLKENKEAKYSKREK 356  
 |||||  
 Db 299 TLTLTD--RGILVQEVNLYVMKATQPODSNTLTCEWGPSTPKMRLIKENOEARVGRQEK 356  
 |||||  
 QY 357 PPMVNLPEAGMOCILSDSGOVLLESNIKVLPTWSPNVCPAPBEPSCKTHCPILLG 416  
 |||||  
 Db 357 VLQVQAPBAGVMOCLLSEGEVMDSKIQVL-----SKGLNQWFLAVVLGS 403  
 |||||  
 QY 417 PSVFL 421  
 |||||  
 Db 404 AFSFL 408  
 |||||

## RESULT 16

RMMST4  
 T-cell surface glycoprotein CD4 precursor - mouse  
 N/Alternate names: T-cell differentiation antigen L3T4; T-cell surface antigen T4/Len 3  
 C/Species: Mus musculus (house mouse)  
 C/Date: 30-Jun-1987 #sequence\_revision 30-Jun-1987 #text\_change 16-Jul-1999  
 C/Accession: A02110; A26038; A39893; A39955; I54564; I65018; A47642  
 R/Tourvilleille, B.; Gorman, S.D.; Field, E.H.; Hunkapiller, T.; Parnes, J.R.  
 Science 234, 610-614, 1986  
 A/Title: Isolation and sequence of L3T4 complementary DNA clones: expression in T cells  
 A/Reference number: A02110; MUID:87018845; PMID:3094146  
 A/Accession: A02110  
 A/Molecule type: mRNA  
 A/Residues: 1-457 <NOU>  
 A/Cross-references: GB:M13816; NID:9192070; PIDN:AAA37267.1; PID:9309112  
 R/Littman, D.R.; Gertner, S.N.  
 Nature 325, 453-455, 1987  
 A/Title: Unusual intron in the immunoglobulin domain of the newly isolated murine CD4 T  
 A/Reference number: A26038; MUID:87115821; PMID:3027575  
 A/Accession: A26038

A: Molecule type: mRNA  
 A: Residues: 1-457 <LIT>  
 A: Cross-references: GB:X04836; NID:G50353; PIDN:CAA28539.1; PID:G50354  
 R: Gorman, S.D.; Tourville, B.; Barnes, J.R.  
 A: Title: Acad. Sci. U.S.A. 84, 7644-7648, 1987  
 A: Title: Structure of the mouse gene encoding CD4 and an unusual transcript in brain.  
 A: Reference number: A39893; MUID:88041159; PMID:2823269  
 A: Accession: A39893  
 A: Molecule type: DNA  
 A: Residues: 1-25, 'E', 27-457 <GOR>  
 A: Cross-references: GB:M17080; GB:J03003; NID:G192515; PIDN:AAA37402.1; PID:G387124  
 R: Maddon, P.J.; Molleaux, S.M.; Maddon, D.E.; Zimmermann, K.A.; Godfrey, M.; Alt, F.W.;  
 Proc. Natl. Acad. Sci. U.S.A. 84, 9155-9159, 1987  
 A: Title: Structure and expression of the human and mouse T4 genes.  
 A: Reference number: A39955; MUID:88097446; PMID:3501122  
 A: Accession: A39955  
 A: Status: nucleic acid sequence not shown; not compared with conceptual translation  
 A: Molecule type: mRNA  
 A: Residues: 25-457 <MAD>  
 A: Note: the cited GenBank accession number, J03564, is not in release 101.0  
 R: Barnes, J.R.; Hunkapiller, T  
 Immunol. Rev. 100, 109-127, 1987  
 A: Title: L3T4 and the immunoglobulin gene superfamily: New relationships between the imm  
 A: Reference number: 154564; MUID:88152875; PMID:3326818  
 A: Accession: 154564  
 A: Status: translated from GB/EMBL/DBJ  
 A: Molecule type: mRNA  
 A: Residues: 1-457 <RRS>  
 A: Cross-references: GB:M36850; NID:G198670; PIDN:AAA39401.1; PID:G198671  
 A: Accession: 169018  
 A: Status: translated from GB/EMBL/DBJ  
 A: Molecule type: DNA  
 A: Residues: 208-318 <RE2>  
 A: Cross-references: GB:M36851; NID:G198672; PIDN:AAA39402.1; PID:G554183  
 R: Claesson, B.U.; Tsagaratos, J.; Kirsbaum, L.; Maddox, J.; Mackay, C.R.; Brandon, M.; M  
 Immunogenetics 23, 129-132, 1986  
 A: Title: The L3T4 antigen in mouse and the sheep equivalent are immunoglobulin-like.  
 A: Reference number: A47642; MUID:86166694; PMID:3082751  
 A: Accession: A47642  
 A: Molecule type: protein  
 A: Residues: 27-43 <CLA>  
 C: Comment: This protein is expressed on most thymocytes, on a subset of mature T-cells  
 C: Genetics:  
 A: Introns: 18/1; 74/1; 128/1; 207/1; 319/1; 386/1; 425/3; 448/2  
 C: Superfamily: T-cell surface glycoprotein CD4; immunoglobulin homology  
 C: Keywords: alternative initiator; duplication; glycoprotein; T-cell; transmembrane pro  
 F: 1-26/Domain: signal sequence #status predicted <SIG>  
 F: 27-457/Product: T-cell surface glycoprotein CD4 #status experimental <MAT>  
 F: 35-114/Domain: immunoglobulin homology <IM1>  
 F: 119-190/Domain: immunoglobulin homology #status atypical <IM2>  
 F: 240-301/Domain: immunoglobulin homology <IM3>  
 F: 241-457/Product: CD4, brain-specific short form #status predicted <BRA>  
 F: 331-372/Domain: immunoglobulin homology <IM4>  
 F: 395-419/Domain: transmembrane #status predicted <TM>  
 F: 440-457/Domain: intracellular #status predicted <INT>  
 F: 447-112 159-188 328-370/Disulfide bonds: #status predicted  
 F: 167, 288, 323, 392/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 29.1%; Score 993.5; DB 1; Length 457;  
 Best Local Similarity 50.9%; Pred. No. 3.6e-51;  
 Matches 217; Conservative 66; Mismatches 120; Indels 23; Gaps 8;

QY 1 MNRGVPRRH-LLVVLQALALPAATQGNKVLGKGDVVELTCTASOKKSIOFHKNSNOI 59  
 DB 1 MCRARISRRRLLLQLLQSLAVTQKTLVIGKEGESALPECSOSKKTIVFTWKFSROR 60  
 QY 60 KILANOG-SFLTQK--PSKINDRADSRSLMDGNFPLIINKLTIKEDSDTYICEVDOK 116  
 DB 61 KILQHGKGVLRIGSGSQF-DRPDSKKGAKGKSPFLIKLMEBSQITICELENNKE 119  
 QY 117 EVQLVLTGELTANSDTHLQGSGLTTLTLES-PPGSSPSVQCRSPGKNIQGGKTLVSQLE 175  
 DB 120 EELWLVFVKTFSPTGSLILQGGSLTLTLDNSKVENPLRECKGKGGKGVVSGKVLSSNLR 179

QY 176 LQSGTWTCTVLQNOKKVBFKIDIVLAFQKASIVYKKEGEVFSFPLAFTVEKLTGS 235  
 DB 180 VQSDPFCNCTVTLQDKQKMGMTLSVGFQSTATAYKSEGESAFEPFLNFAEE--NGW 237  
 QY 236 GELWMAQERASSSKSWITFDPLKKNKESVSKVTDOPKQLQMKKPLHLTLPLQALPOVAGSG 295  
 DB 238 GELWMAKAKKSPQPMWISFSIKKNEVSGQSTDLKQLKMETLPLTLKIQVSLQFAGSG 297  
 QY 296 NLTLAEAKTGKHLQEVNLVYMRATQLOKULTCGEWGPSPKMLSLKENKAKVSKRE 355  
 DB 298 NLTLTLD--KGTILQEVNLVYMRATQLOKULTCGEWGPSPKMLSLKENKAKVSKRE 355  
 QY 356 KPVVNLVPEAGMGMOCLISDSGOVLTLSNIVLPTWSPVPCAPPEPSCKTHCPBLTG 415  
 DB 356 KVVQVNLVPEETGLMWQCLLSBGRKAMDRIQV-----SRVQDTYFLACVLG 402  
 QY 416 GPSVFL 421  
 DB 403 GSFGFL 408

RESULT 17  
 G3MGM  
 IG gamma-3 chain C region, membrane-bound form - mouse  
 C: Species: Mus musculus (house mouse)  
 C: Date: 13-Aug-1986 #sequence\_revision 31-Mar-1991 #text\_change 16-Jul-1999  
 C: Accession: A02156; A02155  
 R: Wells, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blat  
 EMBO J. 3, 2041-2046, 1984  
 A: Title: Structure analysis of the murine IgG3 constant region gene.  
 A: Reference number: A02156; MUID:85027161; PMID:6092053  
 A: Accession: A02156  
 A: Molecule type: DNA  
 A: Residues: 1-398 <WEL>  
 A: Cross-references: GB:J00451; NID:G194392; PIDN:AA859655.1; PID:G194433  
 A: Note: the sequence was determined from the germ-line gene  
 R: Komaromy, M.; Clayton, L.; Rogers, J.; Robertson, S.; Kettman, J.; Wall, R.  
 Nucleic Acids Res. 11, 6775-6785, 1983  
 A: Reference number: A02155; MUID:84041483; PMID:6314258  
 A: Accession: A02155  
 A: Molecule type: DNA  
 A: Residues: 328-332, 'G', 334-341, 'Q', 343-387, 'F', 389-398 <KOM>  
 A: Cross-references: GB:X00688  
 A: Note: the sequence was determined from the germ-line gene  
 C: Genetics:  
 A: Introns: 97/1; 113/1; 223/1; 328/1; 371/3  
 C: Complex: An immunoglobulin heterotrimer subunit consists of two identical light (ka  
 hain disulfide bonds. In some cases, such as Iga and Igm, the subunits associate into 1  
 C: Superfamily: immunoglobulin C region; immunoglobulin homology  
 C: Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglob  
 F: 19-83/Domain: immunoglobulin homology <IM1>  
 F: 97-112/Region: hinge  
 F: 136-305/Domain: immunoglobulin homology <IM2>  
 F: 242-309/Domain: immunoglobulin homology <IM3>  
 F: 346-362/Domain: transmembrane #status predicted <TM>  
 F: 363-398/Domain: intracellular #status predicted <INT>  
 F: 179, 322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 27.5%; Score 938.5; DB 1; Length 398;  
 Best Local Similarity 66.3%; Pred. No. 5.2e-48;  
 Matches 169; Conservative 39; Mismatches 40; Indels 7; Gaps 2;

QY 394 VPCAPPEPKSCDTHNCP--ELIGSPVFLFPPPKDMLISRTPEVTCVVVDSHPDE 451  
 DB 100 IPRKSTPPGS-----SCPENIIGSPVFLFPPPKDMLISRTPEVTCVVVDSHPDE 154  
 QY 452 VKENWYDGVGVNNAKTKPREBOGNSYRVVSVLTVLHQDMLNGEKYKCVSNKALPAPI 511  
 DB 155 VHWVSWFVDNKEVHTAMTQPREADYNSYFRVVSALPIQDQMGKGEKCKVNNKALPAPI 214  
 QY 512 EKTISKAKQGPREEQVYTLPRSDLELNQVSLCLVKGYPEDIANVWESNQPENNYK 571

[illegible]

Db 233 EGMADKXSLTCMIDDFPREDITIVEMQNMNQAPENYKNTQPIKNTNGSVFYFYSKLTNVQSS 292

Qy 595 RMOQGNVFSQSVMEHALNNHYTOKSLSLSPGLQDLDFCAEQDSELDGLMTT 646

Db 293 NWEANTFTCSYVLHSGLNNHHTKSLSHSPGLQDLDFCAEQDSELDGLMTT 344

RESULT 19

CHRB

Ig gamma chain C region - rabbit

C:Species: Oryctolagus cuniculus (domestic rabbit)  
C:Date: 24-Apr-1984 #sequence revision 15-Nov-1984 #text change 16-Jul-1999  
C:Accession: A91749; A90290; A93928; A90245; A94416; A02161  
J:Bernstein, K.E.; Alexander, C.B.; Mage, R.G.  
I:Immunogenetics 18, 387-397, 1983  
A:Title: Nucleotide sequence of a rabbit IgG heavy chain from the recombinant F-1 haplo-  
A:Reference number: A91749; PMID:84030930; PMID:6313520  
A:Accession: A91749

A:Molecule type: mRNA  
A:Residues: 1-323 <BER>  
A:Note: This sequence has the d12 allotypic marker, 104-Thr, and the e14 marker, 185-Th.  
R:Pratt, D.M.; Mole, L.E.  
Biochem. J. 151, 337-349, 1975  
A:Title: Sequence studies on the constant region of the Fd sections of rabbit immunoglo-  
A:Reference number: A90290; PMID:76135469; PMID:1243651  
A:Accession: A90290

A:Molecule type: protein  
A:Residues: 1-47, 'E', 49-71, 'PV', 72-128 <PRA>  
R:Martens, C.L.; Moore, K.W.; Steinmetz, M.; Hood, L.; Knight, K.L.  
Proc. Natl. Acad. Sci. U.S.A. 79, 6018-6022, 1982  
A:Title: Heavy chain genes of rabbit IgG: isolation of a cDNA encoding gamma heavy chain  
A:Reference number: A93928; PMID:83293917; PMID:6193512  
A:Accession: A93928

A:Molecule type: mRNA  
A:Residues: 88-103, 'M', 105-143, 'E', 145-184, 'A', 186, 'E', 188-266 <MAR>  
A:Cross-references: GB:M16426; NID:G16511; PIDN:AAA31289.1; PID:G16512  
A:Note: This sequence has the d11 allotypic marker, 104-Met, and the e15 allotypic mark  
R:Frutcher, R.G.; Jackson, S.A.; Mole, L.E.; Porter, R.R.  
Biochem. J. 116, 249-259, 1970  
A:Title: Sequence studies of the Fd section of the heavy chain of rabbit immunoglobulin  
A:Reference number: A90245; PMID:70110015; PMID:5461106  
A:Accession: A90245

A:Molecule type: protein  
A:Residues: 132-143, 'E', 145-161 <FRU>  
R:Hall, R.L.; Leboyvitz, H.E.; Fellows Jr., R.E.; Delaney, R.  
In Gamma Globulins, Nobel Symp. 3, Kallander, J., ed., 109-127, Almqvist and Wiksell,  
A:Reference number: A94416  
A:Accession: A94416

A:Molecule type: protein  
A:Residues: 129-131, 155-172, 'D', 174-184, 'A', 186, 'E', 188-200, 'D', 202-217, 'E', 219-232, 'Q',  
A:Note: This has the e15 allotypic marker, 185-Ala  
C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kai  
chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1  
C:Superfamily: immunoglobulin C region; immunoglobulin homology  
C:Keywords: duplication; glycoprotein; heterotrimer; immunoglobulin  
F:20-82/Domain: immunoglobulin homology <IM1>  
F:130-199/Domain: immunoglobulin homology <IM2>  
F:236-303/Domain: immunoglobulin homology <IM3>  
F:173/Binding site: carbohydrate (Aam) (covalent) #status predicted

Query Match 26.7%; Score 910, DB 1; Length 323;  
Best Local Similarity 65.7%; Pred. No. 1, 9e-46;  
Matches 174; Conservative 31; Mismatches 44; Indels 16; Gaps 5;

Qy 375 SGQVLESNIKVLPTWSTPVC-----PAPR-----PSCDTHNC--PELLGSPSVF 420

Db 60 SGLYSLSSVSVSTSS--SQPTCVVAHPATNTKVDKYVAPSTCSRP--TCEPPELLGSPSVF 117

Qy 421 LFPKPKDTLMTSRTEPVTCVVVDVSHEDPEVFNNYVDGVENNAKTRPREQYNSTR 480

Db 118 IFPPKPKDTLMTSRTEPVTCVVVDVSHEDPEVFNNYVDGVENNAKTRPREQYNSTR 177

Qy 481 VSVSLTVLHQMVLNGEKYKCKVSNKALPADIEKTIISAKQGPREDQVYTTLPSSRDELTKN 540

Db	178	VYSTLPITHQDWLRGRGEFKCKVHNKALPARIETKITSKARGQPLEPRVYIMGPRELSSR	237
Qy	541	QVSLTCLYKGFPPSDIAVWESNGQBPENNYKTTTPVLDSGSEFLYSKLTVDKSRMOQN	600
Db	238	SVSLTLCMNGFPSPDISVEMENKGRKEDNYKTTPAVLDSGSEFLYNKLSVPTSEMQRGD	297
Qy	601	VFSCSVMEHALNNHTOKSLSPG	625
Db	298	VFTCSVMHEALNNHTOKSLSSRG	322

**RESULT 20**

IG gamma-2a chain C region, membrane-bound form - mouse  
 C:Species: Mus musculus (house mouse)  
 C>Date: 19-Feb-1994 #sequence\_revision 31-Mar-1991 #text\_change 16-Jul-1999  
 C:Accession: A02154; B32657; I57809  
 R:Yamawaki-Kataoka, Y.; Nakai, S.; Miyata, T.; Honjo, T.  
 Proc. Natl. Acad. Sci. U.S.A. 79, 2623-2627, 1982  
 A>Title: Nucleotide sequence of gene segments encoding membrane domains of immunoglobulin  
 A:Reference number: A02154; MUID:82222190; PMID:6283537  
 A:Accession: A02154  
 A:Molecule type: DNA  
 A:Residues: 329-399 <YAM>  
 A:Cross-references: GB:J00471  
 A>Note: The sequence was determined from the germ-line gene  
 R:Yamawaki-Kataoka, Y.; Miyata, T.; Honjo, T.  
 Nucleic Acids Res. 9, 1365-1381, 1981  
 A>Title: The complete nucleotide sequence of mouse immunoglobulin gamma 2 A gene and evolu  
 A:Reference number: A32657; MUID:81198976; PMID:6262729  
 A:Accession: B32657  
 A:Molecule type: DNA  
 A:Residues: 1-329, 'K' <YAZ>  
 R:Hall, B.; Mlicarek, C.  
 Mol. Immunol. 26, 819-826, 1989  
 A>Title: Sequence and polyadenylation site determination of the murine immunoglobulin ga  
 A:Reference number: I57809; MUID:90097953; PMID:2513486  
 A:Accession: I57809  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 373-389 <RES>  
 A:Cross-references: GB:M5032; NID:G194478; PIDN:AAA37919.1; PID:G367217  
 C:Comment: The sequence of residues 1-328 was assumed to be identical with the correspond  
 C:Comment: Cell lines producing IIG conserved two mRNA species for Ig gamma chains. The ma  
 hat it contains an alternative 3' end, encoded in separate exons, that is homologous with  
 C:Genetics:  
 A:Introns: 1/1, 98/1; 114/1; 224/1; 329/1; 372/1  
 C:Complex: An immunoglobulin heterotrimeric subunit consists of two identical light (kappa)  
 chain disulfide bonds. In some cases, such as IGA and IGM, the subunits associate into la  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; heterotrimer; immunoglobulin  
 F:117-206/Domain: immunoglobulin homology <IMM>  
 F:346-353/Domain: transmembrane #status predicted <TM>  
 F:364-359/Domain: intracellular #status predicted <INT>  
 F:180/Binding site: carbohydrate (Asn) (covalent) #status predicted

QY	460	GVEVHNATKTPREEDQYNSYRVVSLVTLVHQQMLNGKEKCKVSNALPAPLEKTSKAK	519
Db	164	NVEVHTAQTQTHREDYNSTLRVVSALPIQHQQMSKEKCKVNNKQDLPAPLEKTSKAK	223
QY	520	GQPREPOVYTLPPSHDELTKQVSLTCLVKGFPSPDAVEMESNGOPENNYKTTPEVLDS	579
Db	224	GSVRAPQYVYVLPPEPEEMTKKQVTLTCWVTDMPEDLYEVMNNNGKLTALNYKNTPEVLDS	283
QY	560	DGSFPLYSKLTLYDKSRMOQGNVFGSSGVNHEALHNHTQSLSLSPQLDDETCAEQDGE	639
Db	284	DGSYMYMSKLRVEKKNQWVERNSYSCSVVHEGLHNHHTTSSFSRTPELDLDVCAEAQDGE	343
QY	640	LDGLWTT	646
Db	344	LDGLWTT	350

## RESULT 21

Ig gamma 2b chain constant region - pig (fragment)  
 C:Species: Sus scrofa domestica (domestic pig)  
 C:Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 21-Jan-2000  
 C:Accession: I47160  
 R:Kacskovics, I.; Sun, J.; Butler, J.E.  
 J. Immunol. 153, 3565-3573, 1994  
 A:Title: Five putative subclones of swine IgG identified from the cDNA sequences of a  
 A:Reference number: I47158; MUID:95015845; PMID:7930579  
 A:Accession: I47160  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-328 <KAC>  
 A:Cross-References: EMBL:U03780; NID:g433125; PIDD:AAA52218.1; PID:g433126  
 C:Genetics:  
 A:Gene: IgG2b  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 F:133-202/Domain: immunoglobulin homology <IM>

## RESULT 22

I47159  
Ig gamma 2a chain constant region - pig (fragment)

C:Species: Sus scrofa domestica (domestic pig)  
 C>Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 21-Jan-2000  
 C/Accession: I47159  
 R/Kaczkovics, I.; Sun, J.; Butler, J.E.  
 J. Immunol. 153, 3565-3573, 1994  
 A>Title: Five putative subclasses of swine IgG identified from the cDNA sequences of a  
 A/Reference number: I47158; MID:95015845; PMID:7930579  
 A/Accession: I47159  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-328 <KAC>  
 A/Cross-references: EMBL:U03779; NID:9433123; PIDN:AAA52220.1; PID:9433124  
 C/Genetics:  
 A/Gene: IGG2a  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 F:133-202/Domain: immunoglobulin homology <IMW>

Query Match 26.4%; Score 901.5; DB 2; Length 328;  
 Best Local Similarity 50.1%; Pred. No. 6.2e-46;  
 Matches 185; Conservative 46; Mismatches 73; Indels 65; Gaps 8;  
 QY 265 RVTQDPRLQWKKLPHLTLPQALPOYAG---SGNLTLEAKTKGLHDEVNLVNRATQ 321  
 16 RDTSGPVALGCLASVFPEPVTVWNSGALSSGVHTFPESVLQPSGLYSLSSMVTVPAS 75  
 QY 322 L-QKNTLCEVWG-P-TSPKMLSLKLEKAKVSKREKPVVNLPEAGMOCILSDSGQVL 379  
 76 LSKSYTCVNNHNPATTKV-----DKRVGTGKTP----- 104  
 DB 380 LSNINIKVLPWTSPVPVCPAPEPESCDKTHTCPELLG-GPSVFLPFPKXDTLMSITREY 438  
 105 -----PSP-----ICPACESGPEVFLFPKPKDTLMSITRPV 138  
 QY 439 TCVVVDVSHDPEVKFNWYDGVGVHNAKTKPEBEQYNSTRYVSVLTVLHODMLNGKEY 498  
 139 TCVVVDVSGNPEVQFSWYDGVGVHNAQTRPKBEQNSYTRVSVLPPIHQDMLNGKEY 198  
 DB 499 KCVSKNALPAPLEKTSKAKGPREQVYTLPPSRDELTKNOVSLTCLVKGYPSPDIAY 558  
 199 KCVKNNMDLPAPITRIISKAKGQTRPEQVYTLPPHAEILSRSKYSTICLVIGYPPDIDV 258  
 QY 559 EWESNQG--PENNYKTPPLDSDGSFELSKLTVDSKRMQGVNFCSMVHEALHNHYT 616  
 259 EWQNGGPEPEGNRTTPPOQDVDTYFLYSKFSVDKASWQGGIFQCAVMHEALHNHYT 318  
 DB 617 QKSLSLSPG 625  
 QY 319 QKISKTRPG 327

RESULT 23  
 147162  
 I: gamma 4 chain constant region - pig (fragment)  
 C/Species: Sus scrofa domestica (domestic pig)  
 C>Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 21-Jan-2000  
 C/Accession: I47162  
 R/Kaczkovics, I.; Sun, J.; Butler, J.E.  
 J. Immunol. 153, 3565-3573, 1994  
 A>Title: Five putative subclasses of swine IgG identified from the cDNA sequences of a  
 A/Reference number: I47158; MID:95015845; PMID:7930579  
 A/Accession: I47162  
 A/Status: preliminary; translated from GB/EMBL/DBJ  
 A/Molecule type: mRNA  
 A/Residues: 1-277 <KAC>  
 A/Cross-references: EMBL:U03782; NID:9433129; PIDN:AAA52220.1; PID:9433130  
 C/Genetics:  
 A/Gene: IGG4  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 F:82-151/Domain: immunoglobulin homology <IMW>

Query Match 26.2%; Score 894; DB 2; Length 277;  
 Best Local Similarity 68.8%; Pred. No. 1.4e-45;  
 Matches 163; Conservative 29; Mismatches 31; Indels 14; Gaps 3;

QY 392 TPVCPAPEPKSCDKTHTCPELLG-GPSVFLPFPKXDTLMSITREYTCVVVDVSHEDP 450  
 DB 51 TPEPCP-----ICPACESGPEVFLFPKPKDTLMSITRPVTCVVVDVSGNPE 99  
 QY 451 EVKFNWYDGVGVHNAKTKPREEQYNSTRYVSVLTVLHODMLNGKEYKCVSKNALPAP 510  
 DB 100 EVQFSRYDGVGVHNAQTRPKBEQNSYTRVSVLPPIHQDMLNGKEYKCVKNNMDLPAP 159  
 QY 511 IKTTSKAKGQPREQVYTLPPSRDELTKNOVSLTCLVKGYPSPDIAYEWESNQG--PEN 568  
 DB 160 IRTIISKAKGQTRPEQVYTLPPTEELSRSKVTLCVLGVFPYPPIDVEMQNGGPEPEPG 219  
 QY 569 NYKTPPVVDSDGSFELSKLTVDSKRMQGVNFCSMVHEALHNHYTQKSLSPG 625  
 DB 220 NYRTTPPOQDVDTYFLYSKLAVDKASWQGGIFQCAVMHEALHNHYTQKSLFTPG 276

RESULT 24  
 G2GP  
 I: gamma-2 chain C region - guinea pig  
 C/Species: Cavia porcellus (guinea pig)  
 C>Date: 07-May-1981 #sequence\_revision 07-May-1981 #text\_change 16-Jul-1999  
 C/Accession: A94553; A90352; A90359; A90384; A90385; A90385; A90385  
 R/Tischmann, T.M.  
 submitted to the Atlas, April 1975  
 A/Reference number: A94553  
 A/Accession: A94553  
 A/Molecule type: protein  
 A/Residues: 1-3 <TRI>  
 R/Birchtein, B.K.; Huseain, Q.Z.; Cebra, J.J.  
 Biochemistry 10, 18-25, 1971  
 A>Title: Structure of heavy chain from strain 13 guinea pig immunoglobulin-G(2). III. Am  
 A/Reference number: A90352; MID:71058471; PMID:5538606  
 A/Accession: A90352  
 A/Molecule type: protein  
 A/Residues: 4-68 <BIR>  
 R/Turner, K.J.; Cebra, J.J.  
 Biochemistry 10, 9-17, 1971  
 A>Title: Structure of heavy chain from strain 13 guinea pig immunoglobulin-G(2). II. Am  
 A/Reference number: A90359; MID:71058486; PMID:5538616  
 A/Accession: A90359  
 A/Molecule type: protein  
 A/Residues: 69-133; 312-329 <TR>  
 R/Tacey, D.E.; Cebra, J.J.  
 Biochemistry 13, 4796-4803, 1974  
 A>Title: Primary structure of the C-H2 homology region from guinea pig IgG2 antibodies.  
 A/Reference number: A90384; MID:75036072; PMID:4429665  
 A/Accession: A90384  
 A/Molecule type: protein  
 A/Residues: 134-226 <TR>  
 R/Tischmann, T.M.; Cebra, J.J.  
 Biochemistry 13, 4804-4811, 1974  
 A>Title: Primary structure of the C-H3 homology region from guinea pig IgG2 antibodies.  
 A/Reference number: A90385; MID:75036073; PMID:4609467  
 A/Accession: A90385  
 A/Molecule type: protein  
 A/Residues: 227-311 <TR>  
 R/Oliveira, B.; Lamm, M.E.  
 Biochemistry 10, 26-31, 1971  
 A>Title: Interchain disulfide bridges of guinea pig gamma-2- immunoglobulin.  
 A/Reference number: A90354; MID:71058474; PMID:4922544  
 A/Contents: annotation; disulfide bonds  
 A/Note: Cys-16 is involved in a heavy-light chain bond  
 A/Note: Cys-105, Cys-107, and Cys-110 form inter-heavy chain bonds  
 C/Comment: This chain was isolated from pooled serum of strain 13 inbred guinea pigs.  
 C/Complex: An immunoglobulin heterodimer subunit consists of two identical light (kappa)  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: duplication; glycoprotein; heterodimer; immunoglobulin  
 F:21-81/Domain: immunoglobulin homology <IM1>  
 F:135-204/Domain: immunoglobulin homology <IM2>  
 F:241-310/Domain: immunoglobulin homology <IM3>

F:28-79/Diulfide bonds: #status experimental  
 F:142-202/Diulfide bonds: #status experimental  
 F:178/Binding site: carbohydrate (asn) (covalent) #status experimental  
 F:248-308/Diulfide bonds: #status experimental

Query Match 25.8%; Score 879.5; DB 1; Length 329;  
 Best Local Similarity 69.2%; Pred. No. 1.2e-44;  
 Matches 164; Conservative 27; Mismatches 39; Indels 7; Gaps 3;

QY 393 PVPCPAPKSCDTHTC--PELLGSPVFLPPPKDTHLSTPEVTCVVDVSHEDP 450  
 D 96 PIRTPPBPCTCPK--CPPEENLGGPSVFLFPKPKDTLMISLTPRTCVVVDSDER 152  
 QY 451 EVKENVYVDGVEVNAKTPREEOYNSTYRVSVLTVLHODMLNGKCKYKSKALPAP 510  
 D 153 EVQGTWFDNKPVNAETKPRVEQYNTFRVESVLPFHODMLNGKCKYKSKALPAP 212  
 QY 511 IEKTSKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNQP--EN 568  
 D 213 IEKTSKAKGAPRMPDVTTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNQP--EN 568  
 QY 569 NYKTPPVLDSGSEFLYSKLTVDKSRKQGNVSCSVMEALNHYTKSLSPG 625  
 D 273 EYKTPPIEDAGSYFLYSKLTVDKSAWDOGTVTYCSVMHEALNHYTKSLSPG 329

## RESULT 25

G2MSBM  
 Ig gamma-2b chain C region, membrane-bound form - mouse

C/Species: Mus musculus (house mouse)

C/Date: 17-Dec-1992 #sequence revision 31-Mar-1991 #text\_change 16-Jul-1999

C/Accession: C02154; A02158; B02157

R/Yamawaki-Katsoka, Y.; Nakai, S.; Miyata, T.; Honjo, T.

A/Title: Nucleotide sequence of gene segments encoding membrane domains of immunoglobulin

A/Reference number: A02154; MUID:82222190; PMID:6283537

A/Accession: C02154

A/Molecule type: DNA

A/Residues: 335-405 <YAM>

A/Cross-References: GB:J00462

R/Rogers, J.; Choi, E.; Sousa, L.; Carter, C.; Word, C.; Kuehl, M.; Eisenberg, D.; Wall, Cell 26, 19-27, 1981

A/Title: Gene segments encoding transmembrane carboxyl termini of immunoglobulin gamma C

A/Reference number: A02158; MUID:82115295; PMID:6799207

A/Accession: A02158

A/Molecule type: DNA

A/Residues: 335-378 <ROG>

A/Note: The translation of the first exon of the membrane-bound segment is given

R/Yamawaki-Katsoka, Y.; Katsoka, T.; Takahashi, N.; Obata, M.; Honjo, T. Nature 283, 786-789, 1980

A/Title: Complete nucleotide sequence of immunoglobulin gamma2b chain gene cloned from r

A/Reference number: A02157; MUID:80120716; PMID:6766534

A/Accession: B02157

A/Content: a allele

A/Accession: B02157

A/Molecule type: DNA

A/Residues: 1-335 <K> <YA2>

C/Comment: The sequence of residues 1-334 was assumed to be identical with the correspond

C/Comment: Cell lines producing IgG contain two mRNA species for Ig gamma chains. The ma

h at it contains an alternative 3' end, encoded in separate exons, that is homologous wit

C/Genetics:

A/Intons: 1/1; 98/1; 120/1; 230/1; 335/1; 378/3

C/Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kap

hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a

C/Superfamily: Immunoglobulin C region; immunoglobulin homology

C/Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglob

F:143-212/Diulfide bonds: #status experimental <IMM>

F:352-369/Diulfide bonds: #status experimental <IMM>

F:370-405/Diulfide bonds: #status experimental <IMM>

F:166/Binding site: carbohydrate (asn) (covalent) #status experimental

Query Match 25.8%; Score 879.5; DB 1; Length 405;  
 Best Local Similarity 47.7%; Pred. No. 1.6e-44;  
 Matches 176; Conservative 54; Mismatches 86; Indels 53; Gaps 6;

QY 285 POALPOVAGSGNTLALAE-----KTGKLEQEVNLVVMARATOLQNLTCCEWGPSPKLM 339  
 D 34 PESVTYVWNSGLSSVHTPPALLOGLYTMSSSVTPSSWTBQYTCVVAHPAS----- 89

QY 340 LSLKLEKKAQVSKREKPVWVNLNPEAGMOCLLSDSGQVLLSNIKVLPTWSTPPCPAP 399  
 D 90 -----STTVQKLEP-----SGPI-----STINCP-- 110

QY 400 EPKSCDTHTC--PELLGSPVFLPPPKDTHLSTPEVTCVVDVSHEDPEVKRWY 457  
 D 111 ---PCKECHCPAPNLGSPVFLFPNINQVLMISLTPKTCVVDVSDDDPVOISWF 167

QY 458 VDGVEVNAKTPREEOYNSTYRVSVLTVLHODMLNGKCKYKSKALPAPLEKTIK 517  
 D 168 VNVVEYTAQTQTHREDYNTIRVSVLTPFHODMLNGKCKYKSKALPAPLEKTIK 227

QY 518 AKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAVWESNQPENNYKTPPVL 577  
 D 228 IKGLVRAPOYYLLPPPAEQLSRKDVSLTCLVGFNPDISVEMTSNGHTEENYKDTAPVL 287

QY 578 DSDGSEFLYSKLTVDKSRKQGNVSCSVMEALNHYTKSLSPGLDDETCAPAD 637  
 D 288 DSDGSEFLYSKLTVDKSRKQGNVSCSVMEALNHYTKSLSPGLDDETCAPAD 347

QY 638 GELDGLMTT 646  
 D 348 GELDGLMTT 356

## RESULT 26

147161  
 Ig gamma 3 chain constant region - pig (fragment)

C/Species: Sus scrofa domestica (domestic pig)

C/Date: 21-Feb-1997 #sequence\_revision 21-Feb-1997 #text\_change 21-Jan-2000

C/Accession: I47161

R/Kacskovics, I.; Sun, J.; Butler, J.E. J. Immunol. 153, 3565-3573, 1994

A/Title: Five putative subclasses of swine IgG identified from the cDNA sequences of a

A/Reference number: I47158; MUID:95015845; PMID:7930579

A/Accession: I47161

A/Status: preliminary; translated from GB/EMBL/DBJ

A/Molecule type: mRNA

A/Residues: 1-328 <KAC>

A/Cross-References: EMBL:U03781; NID:G433127; PIDN:AAA52219.1; PID:G433128

C/Genetics:

A/Genes: IgG3

C/Superfamily: Immunoglobulin C region; immunoglobulin homology

F:133-202/Diulfide bonds: #status experimental <IMM>

Query Match 25.7%; Score 876.5; DB 2; Length 328;  
 Best Local Similarity 49.6%; Pred. No. 1.8e-44;  
 Matches 183; Conservative 44; Mismatches 77; Indels 65; Gaps 8;

QY 265 RVNQPDLQWKKLPLHLTPQALPOVAG--SGNLTALAEKTKGKLEQEVNLVVMARATQ 321

D 16 RDTSGPVNVALGLASSVTPPEVTMTWNSGLTSGVHTFPEVSLDPSGLYSSWTPASS 75

QY 322 L-QKNLTCEWGP-TSEPKMLSLKENKKAQVSKREKPVWVNLNPEAGMOCLLSDSGVL 379

D 76 LSSGSYTCVNNHATTKV-----DKRVGTRKP----- 104

QY 380 LESNIKVLPTWSTPPCPAPBPSCDTHTCPEL-LGSPVFLPPPKDTHLSTPEV 438

D 105 -----PCP-----ICPGCEVAGSPVFLFPKPKDTLMISQTPV 138

QY 439 TCVVVDVSHEDPEVKENWYVDGVEVNAKTPREEOYNSTYRVSVLTVLHODMLNGK 498

D 139 TCVVVDVSKHAEVQSWYVDGVEVNAETKPREEOYNSTYRVSVLTPLHODMLNGK 198

QY 499 KCKVSKNALPAPLEKTIKSKAKGQPREPOVYTLPPSRDELTKNQVSLTCLVKGFYPSDIAV 558

D 199 KCKVNNVDLPAPITRTISKALIGSREPOVYTLPPPAEELSRSKYVTCVIGYPPDIHV 258







A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-472 <PART>  
 A:Cross-references: EMBL:X69797  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: immunoglobulin  
 F:277-346/Domain: immunoglobulin homology <IMM>

Query Match 24.7%; Score 843; DB 2; Length 472;  
 Best Local Similarity 41.2%; Pred. No. 2.7e-42;  
 Matches 213; Conservative 59; Mismatches 141; Indels 104; Gaps 17;

QY 157 SPKGNIGGKTLVSQLEODSGTWCTVILQNGKV-----EFKIDIVLAFOKA 207  
 DB 11 APRG-----VLGVRLOESGSPSLATLLQTLVTCTIGSFLNNYGVWVWQARCKA 61  
 QY 208 SSIYKKEGEVERSFPLATVEKLTGSG---ELMQAERASSSKMTTFLKKEYSVK 264  
 DB 62 -----LEWLGSGYDEDIDYND-VLKSRLSTTKDTSKQVSLT 98  
 QY 265 R---VTODPKLQMGKCLPLHLTLPOALPOVA--GSGNLTLLBAKTKLHQEVNLVYMR 319  
 DB 99 LSTVTTEDTAVYCARVDYDSHAFAYASYPMKPGLLISVLSAST-----TP 146  
 QY 320 TOLQKMLTCEVWGPTSPKMLSLKLENKEAFYKREKRVWV-----LNPBAGMOCLLS 373  
 DB 147 PKVYPLTSC--CGDTSSSIVTLGCLVS---SYMPEVTVTWNSGALTSQVHTFPALQ 199  
 QY 374 DSGVLESNIKV-----LPTWSTPVP-----CPAPEKSCDKTHTC-- 410  
 DB 200 SSGYSLSSSVTVPASTGAGTFCINVAHPASSTKVRKVRPGGPDPC-KKIC---RCPP 254  
 QY 411 PELGSGSVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPEKFNWYDGEVNAKTKP 470  
 DB 255 PELGSGSVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPEKFNWYDGEVNAKTKP 314  
 QY 471 REEYNSTYRVSVVLTVLHODMLNGEKYCKVSNKALPALEKTIISAKGQPREPOVYTL 530  
 DB 315 REEYNSTYRVSVVLTVLHODMLNGEKYCKVSNKALPALEKTIISAKGQPREPOVYTL 374  
 QY 531 PRSDELTKNOVSLTCLVKGFPESDIAMWESNQP--ENNYKTTTPYLDSDGSFFLYSK 588  
 DB 375 APPEELSKSTLSYTCVLTGFPYDIAMWESNQPESDEKYGTTQLDADSGSYFLYSR 434  
 QY 589 LTVDKSRMOGNYFSCVMHEALHNHYTKSLSPG 625  
 DB 435 LRVDKSMQEGDYACVMHEALHNHYTKSLSPG 471

## RESULT 30

G3MGC  
 Ig gamma-3 chain C region, secreted form - mouse  
 C:Species: Mus musculus (house mouse)  
 C:Date: 17-Mar-1987 #sequence\_revision 31-Mar-1991 #text\_change 16-Jul-1999  
 C:Accession: B02156  
 R:Mele, J.A.; Word, C.J.; Rimm, D.; Der-Balan, G.P.; Martinez, H.M.; Tucker, P.W.; Blact  
 EMBL J. 3, 2041-2046, 1984  
 A:Title: Structure analysis of the murine IgG3 constant region gene.  
 A:Reference number: A02156; MUID:85027161; PMID:6092053  
 A:Molecule type: DNA  
 A:Residues: 1-329 <MEL>  
 A:Cross-references: GB:J00451  
 A:Note: the sequence was determined from the germline gene  
 C:Genetics:  
 A:Insertions: 97/1; 113/1; 223/1  
 C:Complex: An immunoglobulin heterotrimer subunit consists of two identical light (kappa) chain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1d C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: alternative splicing; duplication; glycoprotein; heterotrimer; immunoglobulin  
 F:19-83/Domain: immunoglobulin homology <IM1>  
 F:97-112/Region: hinge  
 F:136-205/Domain: immunoglobulin homology <IM2>

F:242-309/Domain: immunoglobulin homology <IM3>  
 F:179,322/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 24.6%; Score 840.5; DB 1; Length 329;  
 Best Local Similarity 64.5%; Pred. No. 2.4e-42;  
 Matches 151; Conservative 37; Mismatches 39; Indels 7; Gaps 2;

QY 394 VPCPAPKSCDKTHTC--ELLGSPSVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPE 451  
 DB 100 IPRDSTPSS-----SCPBNILIGSPVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPD 154  
 QY 452 VKENWYDGEVNAKTKPREEYNSTYRVSVVLTVLHODMLNGEKYCKVSNKALPAPI 511  
 DB 155 VHSWVFYDNKEVHTAWTOPREAGYNSTFRVVSALPIQHDMMRGKEKCKVNNALPAPI 214  
 QY 512 EKTIISAKGQPREPOVYTLPRSDELTKNOVSLTCLVKGFPESDIAMWESNQPENNYK 571  
 DB 215 ERITISKEGKAQTPQVTVTPPRBQMSKKVSLCLVTNPFSEASIVEMRNGELEQDYK 274  
 QY 572 TTPPYLDSDGSFFLYSKLTVDKSRMOGNYFSCVMHEALHNHYTKSLSPG 625  
 DB 275 NTPPYLDSDGSFFLYSKLTVDTDSMGLDGLFTCSVMHEALHNHYTKSLSPG 328

## RESULT 31

C30554  
 Ig heavy chain C region - sheep (fragment)  
 C:Species: Ovis orientalis aries, Ovis aries (domestic sheep)  
 C:Date: 03-Mar-1989 #sequence\_revision 03-Mar-1989 #text\_change 21-Jan-2000  
 C:Accession: C30554  
 R:Foley, R.C.; Beh, K.J.  
 J. Immunol. 142, 708-711, 1989  
 A:Title: Isolation and sequence of sheep Ig H and L chain cDNA.  
 A:Reference number: A30554; MUID:89093962; PMID:2492052  
 A:Accession: C30554  
 A:Molecule type: mRNA  
 A:Status: preliminary; not compared with conceptual translation  
 A:Residues: 1-308 <POL>  
 C:Superfamily: immunoglobulin C region; immunoglobulin homology  
 C:Keywords: heterotrimer; immunoglobulin  
 F:113-182/Domain: immunoglobulin homology <IMM>

Query Match 24.5%; Score 836; DB 2; Length 308;  
 Best Local Similarity 49.1%; Pred. No. 4e-42;

Matches 173; Conservative 39; Mismatches 76; Indels 64; Gaps 7;

QY 284 LPQALPOYAGSGNLT-----ALEAKTKLHQEVNLVWRATQLOKMLTCEVWGPTSPK 337  
 DB 10 MPEVTVTWNSGALTSQVHTFPALQSSGLYSVTVPASTGAGTFCINVAHPAS-- 67  
 QY 338 LMLSLKLENKEAFYKREKRVWVNLNPBAGMOCLSDSGVLLSNIKVLPTWSTPVP 397  
 DB 68 -----STVYDKRVEP-----GCPDPC- 83

QY 398 APPEKSCDKTHTC--PELIGSPSVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPEKFN 455  
 DB 84 -----KPC-----RCPPELPGSPSVFLFPKPKDQTLMSRTPEYTCVNVVSHEDPEKFN 135

QY 456 WYDGEVNAKTKPREEYNSTYRVSVVLTVLHODMLNGEKYCKVSNKALPALEKTI 515  
 DB 136 WYDGEVNAKTKPREEYNSTYRVSVVLTVLHODMLNGEKYCKVSNKALPALEKTI 195

QY 516 SKAGQPREPOVYTLPRSDELTKNOVSLTCLVKGFPESDIAMWESNQP--ENNYKTT 573  
 DB 196 SKAGQPREPOVYTLPRSDELTKNOVSLTCLVKGFPESDIAMWESNQPESDEKYGTT 255

QY 574 PLYLDSDGSFFLYSKLTVDKSRMOGNYFSCVMHEALHNHYTKSLSPG 625  
 DB 256 TSQLDADSGFFLYSKLTVDKSRMOGNYFSCVMHEALHNHYTKSLSPG 307

## RESULT 32

PC4436

monoclonal antibody 13-1 heavy chain - mouse (fragment)  
 C/Species: Mus musculus (house mouse)  
 C/Date: 04-Feb-1998 #sequence\_revision 13-Mar-1998 #text\_change 21-Jan-2000  
 C/Accession: PC4436  
 R/Akashi, S.; Kato, K.; Torizawa, T.; Dohmae, N.; Yamaguchi, H.; Kamachi, M.; Harada, A.  
 Biochem. Biophys. Res. Commun. 240, 566-572, 1997  
 A/Title: Structural characterization of mouse monoclonal antibody 13-1 against a porphyrin  
 A/Reference number: JCS810; MUID:98063277; PMID:9398605  
 A/Accession: PC4436  
 A/Molecule type: protein  
 A/Residues: 1-444 <AKA>  
 C/Comment: This catalytic antibody has peroxidase oxidase activity. It is directed again  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 F/251-320/Domain: immunoglobulin homology <IMM>  
 F/22/Disulfide bonds: Interchain (to 98) #status predicted  
 F/99/Disulfide bonds: Interchain (to 109) #status predicted

Query Match 24.3%; Score 830.5; DB 2; Length 444;  
 Best Local Similarity 57.0%; Pred. No. 1.4e-41;  
 Matches 150; Conservative 44; Mismatches 49; Indels 20; Gaps 4;

QY 380 LESNIV-LPTW-STVPCPAPRPSKCDKTH-----TCPELLGAPSVFLF 422  
 Db 184 LSSSVTPSSSTWPSSETVTCNVAPASSTKVDKIVPRDGCXPCICTVPEV---SSVFIF 240  
 QY 423 PPRPKDTLMSRPEVTCVVVDVSHEDPEVKFNNWYDGVVHNAAKTRPREQYNSTYRVY 482  
 Db 241 PPRPKDTLITLTPKVCVVVDISKDDPEVFQFWMFDDVVAQTQPREQYNSTYRVY 300  
 QY 483 SVTLVHODMLNGEKYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNOY 542  
 Db 301 SEIPIHODMLNGEKYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDELTKNOY 360  
 QY 543 SLTCLVKGFPYSDIAVEMESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNV 602  
 Db 361 SLTCLVKGFPYSDIAVEMESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRWQQGNV 420  
 QY 603 SCGVMEHALNHNHTQKSLSPG 625  
 Db 421 TCSVLHGLHNNHTKSLSPG 443

RESULT 33  
 PS0017  
 Ig gamma-1 chain C region - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 16-Jul-1999  
 C/Accession: PS0017; C25941  
 R/Bruggemann, M.  
 Gene 74, 473-482, 1988  
 A/Title: Evolution of the rat immunoglobulin gamma heavy-chain gene family.  
 A/Reference number: PS0017; MUID:89232738; PMID:3149946  
 A/Accession: PS0017  
 A/Molecule type: DNA  
 A/Residues: 1-326 <BRU>  
 R/Bruggemann, M.; Free, J.; Diamond, A.; Howard, J.; Cobbold, S.; Waldmann, H.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 6075-6079, 1986  
 A/Title: Immunoglobulin heavy chain locus of the rat: striking homology to mouse antio  
 A/Reference number: A25941; MUID:86287397; PMID:3016742  
 A/Accession: C25941  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 220-326 <BR2>  
 C/Genetics:  
 A/Intons: 98/1; 113/1; 220/1  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: immunoglobulin  
 F/20-84/Domain: immunoglobulin homology <IMM>

Query Match 24.3%; Score 829; DB 2; Length 326;  
 Best Local Similarity 57.3%; Pred. No. 1.1e-41;  
 Matches 153; Conservative 42; Mismatches 56; Indels 16; Gaps 4;

QY 375 SGQVLESNIKV-LPTW-STVPCPAPRPSKCDKTH-----PS 418  
 Db 59 SGLYTLTSSV-TPSSSTWPSSETVTCNVAPASSTKVDKIVPRDGCXPCICTGSEVSS 118  
 QY 419 VFLPFPKDTLMSRPEVTCVVVDVSHEDPEVKFNNWYDGVVHNAAKTRPREQYNST 478  
 Db 119 VFLPFPKDTLITLTPKVCVVVDISGDDPEVHFMSWVDVVAQTQPREQYNST 178  
 QY 479 YRVVSVTLVHODMLNGEKYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDEL 538  
 Db 179 FRSVSELPIHODMLNGEKYCKVSNKALPAPIEKTISKAKGQPREPOVYTLPPSRDEL 238  
 QY 539 KNQVSLTCLVKGFPYSDIAVEMESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRW 598  
 Db 239 QNEVSLTCLVKGFPYSDIAVEMESNGQPENNYKTTPPVLDSDGSFFLYSKLTVDKSRW 298  
 QY 599 GNVFSCVMEHALNHNHTQKSLSPG 625  
 Db 299 GNTFTCSVLHGLHNNHTKSLSPG 325

RESULT 34  
 PS0018  
 Ig gamma-2b chain C region - rat  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 07-Jun-1990 #sequence\_revision 07-Jun-1990 #text\_change 16-Jul-1999  
 C/Accession: PS0018; B25941  
 R/Bruggemann, M.  
 Gene 74, 473-482, 1988  
 A/Title: Evolution of the rat immunoglobulin gamma heavy-chain gene family.  
 A/Reference number: PS0017; MUID:89232738; PMID:3149946  
 A/Accession: PS0018  
 A/Molecule type: DNA  
 A/Residues: 1-333 <BRU>  
 R/Bruggemann, M.; Free, J.; Diamond, A.; Howard, J.; Cobbold, S.; Waldmann, H.  
 Proc. Natl. Acad. Sci. U.S.A. 83, 6075-6079, 1986  
 A/Title: Immunoglobulin heavy chain locus of the rat: striking homology to mouse antio  
 A/Reference number: A25941; MUID:86287397; PMID:3016742  
 A/Accession: B25941  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 227-333 <BR2>  
 C/Genetics:  
 A/Intons: 96/1; 117/1; 227/1  
 C/Superfamily: immunoglobulin C region; immunoglobulin homology  
 C/Keywords: immunoglobulin  
 F/20-82/Domain: immunoglobulin homology <IMM>

Query Match 24.2%; Score 826.5; DB 2; Length 333;  
 Best Local Similarity 57.1%; Pred. No. 1.6e-41;  
 Matches 157; Conservative 35; Mismatches 58; Indels 25; Gaps 4;

QY 375 SGQVLESNIKV-LPTW-STVPCPAPRPSKCDKTH-----CDKTHTC-- 410  
 Db 59 SGLYTLTSSV-TPSSSTWPSSETVTCNVAPASSTKVDKIVPRDGCXPCICTGSEVSS 117  
 QY 411 PELLGSPVFLPFPKDTLMSRPEVTCVVVDVSHEDPEVKFNNWYDGVVHNAAKTRP 470  
 Db 118 PELLGSPVFLPFPKDTLMSRPEVTCVVVDVSHEDPEVKFNNWYDGVVHNAAKTRP 177  
 QY 471 REQYNSTYRVVSVTLVHODMLNGEKYCKVSNKALPAPIEKTISKAKGQPREPOVYTL 530  
 Db 178 REQYNSTYRVVSVTLVHODMLNGEKYCKVSNKALPAPIEKTISKAKGQPREPOVYTL 237  
 QY 531 PPSRDELTKNOYSLTCLVKGFPYSDIAVEMESNGQPENNYKTTPPVLDSDGSFFLYSKLT 590  
 Db 238 GPPTDELTKNOYSLTCLVKGFPYSDIAVEMESNGQPENNYKTTPPVLDSDGSFFLYSKLT 297  
 QY 591 VDKSRWQQGNVPSGVMEHALNHNHTQKSLSPG 625  
 Db 298 VDKSRWQQGNVPSGVMEHALNHNHTQKSLSPG 332

## RESULT 35

S37483

Ig gamma-2a chain - mouse

C/Species: Mus musculus (house mouse)

C/Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 23-Jul-1999

C/Accession: S37483

R/Duncan, F.F.D.

Submitted to the EMBL Data Library, February 1993

A/Reference number: S37483

A/Accession: S37483

A/Status: preliminary

A/Molecule type: mRNA

A/Residues: 1-469 &lt;DUC&gt;

A/Cross-references: EMBL:X70423; NID:g406252; PIDN:CAA49868.1; PID:g406253

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: immunoglobulin

F/276-345/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 24.1%; Score 823.5; DB 2; Length 469;  
 Best Local Similarity 35.0%; Pred. No. 3.8e-41;  
 Matches 213; Conservative 63; Mismatches 150; Indels 183; Gaps 19;

```

Qy 30 LGKGGDYELTCTASQKSIQFHMKNQKILGNQ-----SFLTKGPKLNDRADSR 83
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 30 LVKGGASVYKISKAGSYFTDY---IMVKKQKPGQGLKIGWIPASGNTKYNENFKG 86
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 84 RSL---WQGNFPLIKLKIEDSDTYICEVEDQKEBQVLVFGLTANSHTLQ--GQS 138
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 87 ATLVTDTSSSTAYWQLSLTSEDTAVVFC-----ADAMGATAT---LLDYWQG 132
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 139 LTLTLESPGSPSPVQCRSPKKNIQGKTLVSQLELQDSGTCTVLQOKKVEFKID 198
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 133 TTLTVSSAKKTAPSVPLAP---VCGDPT-----LTDYWQG 158
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 199 IVTLAFOKASSIVYKKEGQVEFSFPLAFTVEKLTGSGELMWQERASSSKWTFDLKN 258
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 159 -----GSSVTL---GCLVKGYPF---EPVT---LTMNSGSLSSG----- 188
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 259 KEVSVKRTQDPKQMGKKLPLHLTLPLQALPQ--YAGSGNLTALAEATGTLHGVNLVY 316
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 189 -----VH--TFPVLVQLSDLYTLSSVT-----V 209
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 317 MRATOLQNLTCFVWPTSPKMLSLKLENKAKVSKREKPVVWLNPEAGMOCILSDSG 376
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 210 TTSSTWPSGSLTCVNAHPSS-----STKYDK----- 234
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 377 QVLESNIKVLPTWSTPVPAPAPKSCDKTHTCPELGGPSVFLFPKPKDTLMISRT 436
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 235 -----KIEPRGPTIKCP---PKC-----PAPNLLGSPSVFIFPKIKIDVLMISLP 279
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 437 EVTCVVVDVSHEDPEVKFNWYVDGVEFNATKPREEDQNTSYRVSVLTVLHQDWLNGK 496
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 280 IVTCVVVDVSDDDPVOISWFMVNNVETATQTHREDYNTLTVVSLALPIODHMMWSGK 339
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 497 EYKCKSVNKALPAPIEKTISKAKGPREPOVYTLPPSRDELTKQVSLTCLVKKFYPSDI 556
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 340 EFKCKVNNKDLPAPIEKTISKAKSVRAPOVYVLPPEEETKQVTLTCVNTDFMPEDI 399
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 557 AVESWNGQPPNNYKTPPVLDSDGSPFLYSKLTIVDSKRMQGNVFCSVVHGLAHNYT 616
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 400 YVEWNTNNGKTELNVKTEPVLDSGSYFMYSKLRKKNKWNVERNSYSGSVVHGLAHNHT 459
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 617 QKSLSLSPG 625
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 460 TKSFSPRTPG 468
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

## RESULT 36

S00847

Ig gamma-2c chain C region - rat (fragment)

C/Species: Rattus norvegicus (Norway rat)

C/Date: 01-Dec-1989 #sequence\_revision 01-Dec-1989 #text\_change 23-Jul-1999

C/Accession: S00847

R/Brueggemann, M.; Delmastro-Galfre, P.; Waldmann, H.; Calabi, F.

Eur. J. Immunol. 18, 317-319, 1988

A/Title: Sequence of a rat immunoglobulin gamma-2c heavy chain constant region cDNA: ex

A/Reference number: S00847; MUID:88166903; PMID:3127222

A/Accession: S00847

A/Molecule type: mRNA

A/Residues: 1-329 &lt;BRU&gt;

A/Cross-references: EMBL:X07189; NID:g57602; PIDN:CAA30169.1; PID:g663228

C/Superfamily: immunoglobulin C region; immunoglobulin homology

C/Keywords: immunoglobulin

F/20-84/Domain: immunoglobulin homology &lt;IMM&gt;

Query Match 24.1%; Score 822.5; DB 2; Length 329;  
 Best Local Similarity 57.0%; Pred. No. 2.7e-41;  
 Matches 154; Conservative 47; Mismatches 50; Indels 19; Gaps 4;

```

Qy 375 SGQVLESNIKV-LPTWST-PVPC-----PA-----PEPKCKTHTCPELGG 415
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 59 SGLVTLSSVTVPSSTWSSQVTCVVAHPATKSLIKRIEPRRPKPPPTDICSDDNLG 118
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 416 GPSVFLPPPKPKDTLMISRTPEVTCVVDVSHEDPEVKFNWYVDGVEFNHAKTKPREQY 475
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 119 RPSVFIEPPPKKIDLTLPKVTCTVVDVSEEPDVQFSWFDNVAVFTAQTPHEQL 178
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 476 NSTRVSVTLTVLHQDWLNGEKYCKVSNKALPAPIEKTISKAKGPREPOVYTLPPSRD 535
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 179 NGTRVSTLHIGQDMSSGKEFKCKVNNKDLSPDKITISKRGKAKTPOVITIPPRE 238
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 536 ELTKNQVSLTCLVKGFYPSDIAVWESNGQPPNNYKTPPVLDSDGSPFLYSKLTVDKSR 595
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 239 QMSKNKSLTCLVWTSFSPASISVEMENGELEDQYKNTLPLVLDSESYFLYSKLSVDTDS 298
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Qy 596 WQGNVFCSVVHGLAHNHTQKSLSPG 625
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 299 WMRGDIYTCVSVHGLAHNHTQKSLSPG 328
      |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

```

## RESULT 37

GIMS

Ig gamma-1 chain C region, secreted form - mouse

C/Species: Mus musculus (house mouse)

C/Date: 31-Mar-1980 #sequence\_revision 24-Sep-1981 #text\_change 16-Jul-1999

C/Accession: A02159; A26234; A26236

R/Honjo, T.; Obata, M.; Yamawaki-Kataoka, Y.; Kataoka, T.; Kawakami, T.; Takahashi, N.;

Cell 18, 559-568, 1979

A/Title: Cloning and complete nucleotide sequence of mouse immunoglobulin gamma1 chain

A/Reference number: A02159; MUID:80045036; PMID:115593

A/Accession: A02159

A/Molecule type: DNA

A/Residues: 1-324 &lt;HON&gt;

A/Cross-references: GB:V00453

A/Note: the sequence was determined from the germline gene

A/Note: Lys-324 is removed posttranslationally

R/Obata, M.; Yamawaki-Kataoka, Y.; Takahashi, N.; Kataoka, T.; Shimizu, A.; Mano, Y.; S

Gene 9, 87-97, 1980

A/Title: Immunoglobulin gamma-1 heavy chain gene: structural gene sequences cloned in a

A/Reference number: A26234; MUID:80202559; PMID:6769752

A/Accession: A26234

A/Contents: MOPC 31C

A/Molecule type: mRNA

A/Residues: 76-324 &lt;OBA&gt;

A/Cross-references: GB:V00775; NID:g51652; PIDN:CAA24153.1; PID:g51653

R/Rogers, U.; Clarke, P.; Salter, W.

Nucleic Acids Res. 6, 3305-3321, 1979

A/Title: Sequence analysis of cloned cDNA encoding part of an immunoglobulin heavy chain

A/Reference number: A26236; MUID:80012837; PMID:113776

A/Contents: MOPC 21

A/Accession: A26236

A/Molecule type: mRNA

A/Residues: 170-275, 'D', 277, 'D', 279-322 &lt;ROG&gt;

A/Cross-references: GB:V00795; NID:g51830; PIDN:CAA24176.1; PID:g780265

R/Adetunbo, K.

J. Biol. Chem. 253, 6068-6075, 1978

A>Title: Evolution of immunoglobulin subclasses. Primary structure of a murine myeloma g  
A/Reference number: A26237; MUID:76242288; PMID:98524  
A/Contents: annotation; MOPC 21  
A/Note: this is the final paper in a series reporting the protein sequence, the disulfid  
A/Note: there are a number of differences from the sequence shown  
C/Genetics:  
A/Introns: 1/1; 98/1; 111/1; 218/1  
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (kap  
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a  
C/Superfamily: immunoglobulin C region; immunoglobulin homology  
C/Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob  
F:20-84/Domain: immunoglobulin homology <IM1>  
F:98-110/Region: hinge  
F:131-200/Domain: immunoglobulin homology <IM2>  
F:237-304/Domain: immunoglobulin homology <IM3>  
F:27-82,138-198,244-302/Disulfide bonds: #status experimental  
F:102/Disulfide bonds: interchain (to light chain) #status experimental  
F:104,107,109/Disulfide bonds: interchain (to heavy chain) #status experimental  
F:174/Binding site: carbohydrate (Asn) (covalent) #status experimental

Query Match 23.9%; Score 814.5; DB 1; Length 324;  
Best Local Similarity 53.9%; Pred. No. 7,9e-41;  
Matches 146; Conservative 51; Mismatches 47; Indels 27; Gaps 3;

Qy 379 LLESNIKVLTWSTPRPCAP-EPKSGDKTH-----TCPELL 414  
Db 56 VLQSDLYLSSSVTVBSSPRPSTVTCVNAHPASSTKVDKVIYRQCKPCCTVPEV- 114  
Qy 415 GGSVFLFPPKPKDTLMISRTPEVTCVVDVSHEDDEVKFNWYVDGEVNAKTPRERQ 474  
Db 115 --SSVFIFPPKPKDVLITLTPKVTCTVVDISKDEPEVQSWVDVENVTAQTQPREQ 172  
Qy 475 YNSTYRVSVLTVLHODMLNGKYYKCKVSKALPAPLEKITSAPAKQPREPQYTTLPSP 534  
Db 173 FNSTFVSSELPIPHODMLNGKKEFKCKVNSAAPPAPLEKITSKTPKAPQYTTLPSP 232  
Qy 535 DELTKNOVSLTCLVKGYSPDIVEMESNGCPENNYKTPPVLDGSEFLYSKLVNDS 594  
Db 233 EQPAKDVSLTCTITDPFEDITVEWQNPAPENNYKNTPIVNTGSIFFVSKLVNQS 292  
Qy 595 RMOGNVFCGVNHEALHNYTKSLSPG 625  
Db 293 NMEAGNFTCSVLHEGHNNHTKSLSHSG 323

RESULT 38  
G2MS11  
Ig gamma-2b chain - mouse  
C/Species: Mus musculus (house mouse)  
C/Date: 31-Mar-1980 #sequence revision 01-Dec-2000 #text change 01-Dec-2000  
C/Accession: S25057; A02157; A26232; A26233; A53598  
R/Fischer, R.; Voss, A.; Nierbach, M.; Munziker, W.; Hirsch, H.J.; Kreuzaler, F.  
Submitted to the EMBL Data Library, July 1992  
A/Description: Production of a Tobacco mosaic virus (TMV) inactivating neotop specific n  
A/Reference number: S25057  
A/Accession: S25057  
A/Status: preliminary  
A/Molecule type: mRNA  
A/Residues: 1-474 <FIS>  
A/Cross-references: EMBL:X67210; NID:G54826; PIDN:CAA47649.1; PID:G54827  
R/Yamawaki-Kataoka, Y.; Kataoka, T.; Takahashi, N.; Odate, M.; Honjo, T.  
Nature 283, 786-789, 1980  
A/Title: Complete nucleotide sequence of immunoglobulin gamma2b chain gene cloned from m  
A/Reference number: A02157; MUID:80120716; PMID:6766534  
A/Contents: a allele  
A/Accession: A02157  
A/Molecule type: DNA  
A/Residues: 138-161, 'L', 163-189, 'FP', 193-474 <YAM>  
A/Cross-references: GB:J00461  
A/Note: the sequence was determined from the germline gene  
R/Tucker, P.W.; Marcu, K.B.; Slightom, J.L.; Blattner, F.R.  
Science 206, 1299-1303, 1979  
A/Title: Structure of the constant and 3' untranslated regions of the murine gamma2b hea

A/Reference number: A26235; MUID:80081501; PMID:117548  
A/Contents: MOPC 11  
A/Accession: A26235  
A/Molecule type: mRNA  
A/Residues: 138-172, 'P', 174-189, 'FP', 193-376, 'T', 378-474 <TUI>  
A/Note: Lys-474 is probably removed posttranslationally  
R/Tucker, P.W.; Marcu, K.B.; Newell, N.; Richards, J.; Blattner, F.R.  
Science 206, 1303-1306, 1979  
A/Title: Sequence of the cloned gene for the constant region of murine gamma2b immunogl  
A/Reference number: A26232; MUID:80081502; PMID:117549  
A/Accession: A26232  
A/Molecule type: DNA  
A/Residues: 138-172, 'P', 174-189, 'FP', 193-376, 'T', 378-474 <TUI>  
R/Ollo, R.; Rougeon, F.  
Nature 286, 761-763, 1982  
A/Title: Mouse immunoglobulin allotypes: post-duplication divergence of gamma2a and gam  
A/Reference number: A26233; MUID:82173203; PMID:6803173  
A/Contents: b allele  
A/Accession: A26233  
A/Molecule type: DNA  
A/Residues: 138-161, 'L', 163-189, 'FP', 193-300, 'R', 302-331, 'N', 333-437, 'DI', 440-474 <OLL>  
A/Cross-references: GB:J00461  
R/Kim, H.; Yamaguchi, Y.; Masuda, K.; Matsunaga, C.; Yamamoto, K.; Irimura, T.; Takahashi  
J. Biol. Chem. 269, 12345-12350, 1994  
A/Title: O-glycosylation in hinge region of mouse immunoglobulin G2b.  
A/Reference number: A53598; MUID:94216359; PMID:7512967  
A/Accession: A53598  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 234-251 <KIM>  
C/Comment: The a allele sequence is shown.  
C/Genetics:  
A/Introns: 138/1; 236/1; 258/1; 368/1  
C/Complex: An immunoglobulin heterotetramer subunit consists of two identical light (ka  
hain disulfide bonds. In some cases, such as IgA and IgM, the subunits associate into 1a  
C/Superfamily: immunoglobulin C region; immunoglobulin homology  
C/Keywords: alternative splicing; duplication; glycoprotein; heterotetramer; immunoglob  
F:157-222/Domain: immunoglobulin homology <IM1>  
F:236-257/Region: hinge  
F:281-350/Domain: immunoglobulin homology <IM2>  
F:387-454/Domain: immunoglobulin homology <IM3>  
F:152/Disulfide bonds: interchain (to light chain) #status predicted  
F:164-220,288-348,394-452/Disulfide bonds: #status predicted  
F:247,250,255,256/Disulfide bonds: interchain (to heavy chain) #status predicted  
F:324/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 23.6%; Score 807; DB 1; Length 474;  
Best Local Similarity 33.9%; Pred. No. 3,6e-40;  
Matches 205; Conservative 73; Mismatches 149; Indels 178; Gaps 17;

Qy 34 GPTVELTCTASQKSIQF--HKNSNOIKILNOC---SFL--TKGPKLMDRADSRSS 85  
Db 34 GASVKNSCVASYTFTTYVMW----VKQKGGGIEWIGYINPNNDGTFKFEKFGKAT 88  
Qy 86 LW---DQGNFPLINKLKIEDSDTYICEVEDQKEVQLVFGLTANSDFHLQGGSLTLT 142  
Db 89 LTSKDSNTAYWELSLTSEDSAYVYCARD-----YDYDFATYGGGSLVLT 134  
Qy 143 LSPSPSPSVOCSPRGKNIQGGKTLVSQLELDSGTGTCTVLONQKVEFKIDIVVL 202  
Db 135 VSAAKTTPSVYPLAPGCGDITGSSVTS----- 162  
Qy 203 APOKASISYIKKEGEVSEFPLATVEKLTSGELMQAESAASSKSWITDLYKKEVS 262  
Db 163 -----GCLVKKYFPPESVTVT-----MNSGSLSSS----- 186  
Qy 263 VKRVTQDPKLGKGLPLHLTLPOALPOYAGSNLTALFAKTKLHGEVNLVVMKATOL 322  
Db 187 -----VH--TLSQLAQ---SGLYTMS-----SSVTPGSTVP 214  
Qy 323 OKNLTCFVWGPSPKMLSLKLENKEAKYSKREKPVVNLNPAQMWQCULSDSGQVLLS 382  
Db 215 SQTIVTCSVNAHPAS-----STVYDKKLEP-----SGPI----- 241